```
CM of: US-09-126-945B-2_COPY_21_29 to: GenEmbl:*   out_format : pto

Date: Sep 14, 2001  9:39 PM

About: Results were produced by the GenCore software, version 4.5.
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL frame+_p2n.model -DEV xlp
 -Q "/cgn2_1/USPT_output/09172001/tmp/us_1155_1_1/1/US-09-126-945B-2_COPY_21_29.seq" -DB
 -DB=GenEmbl -QFMT=fastap -SUFFIX=olip2n.rge  GAPOP 4.560
 -GAPEXT 0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT 0.000
 -QGAPOP 4.500 -QGAPEXT=0.050 -XGAPOP=60.000  XGAPEXT 60.000
 -YGAPOP 6.000 -YGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT 60.000
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX olip2
 -FRAME=human40.cdi -LIST=45 -DOCALIGN=200  THR_SCORE=quality
 -THR_MIN 1 -ALIGN 15  MODE LOCAL  -HTMPT pto  NORM ext  MINGEN 0
 -MAXLEN=2000000000 -USER=US09126945_ACGN1_1_7594 -NCPU 6
 -ICPU 4 -LONGLOG -NO_XLPXY -WAIT -THREADS 1

Search information block:
Query: US-09-126-945B-2_COPY_21_29
Query length: 9
Database: GenEmbl:*
Database sequences: 1344157
Database length: -856060004
Search time (sec): 6503.700000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000


score_list:
Sequence                 Strd Orig    QScore    BScore  Len  ! Documentation ...
qb_patl:AX048148          +     9.00   161.93   0.8782    1894  ! AX048148 Sequence 1 from Pate
qb_pr4:AF071538           +     9.00   161.93   0.8782    1894  ! AF071538 Homo sapiens Ets tra
qb_pr1:AB031549           +     9.00   161.92   0.8796    1898  ! AB031549 Homo sapiens PSE mRNA
qb_patl:AX048159          +     9.00   159.65   1.81      3317  ! AX048159 Sequence 12 from Pat
qb_pr6:AL157372           -     9.00   136.32   23.44     159668  ! AL157372 Human DNA sequence
qb_pr8:HS11B8F            +     7.00   134.93   28.02     288  ! Z59914 H.sapiens CpG island DNA
qb_in3:HSZ99661           +     7.00   132.76   27.20     417  ! Z99661 Heterakis sp. 28S rRNA
qb_stsl:CNS06RLJ          +     7.00   127.15   76.04     1089  ! AL412109 T7 end of clone AW0AA
qb_pl3:PSP276438          +     7.00   122.76   133.52    2306  ! AJ276438 Pirellula sp. B2 mRNA
qb_pl1:AF016864           +     7.00   122.44   139.09    2435  ! AF016864 Orpinomyces sp. PC-2
qb_patl:AR049614          -     7.00   121.06   166.08    3084  ! AR049614 Sequence 3 from paten
qb_vil:AF075255           -     7.00   113.45   440.42    11311  ! AF075255 Venezuelan equine en
qb_vil:AF075259           -     7.00   113.41   442.58    11385  ! AF075259 Venezuelan equine en
qb_pr5:AF156545           +     7.00   107.65   926.31    30461  ! AF156545 Homo sapiens alpha-1
qb_patl:AR065852          -     7.00   107.33   965.88    32207  ! AR065852 Sequence 20 from pat
qb_in3:CELK10B3           -     7.00   107.28   971.65    32464  ! U49941 Caenorhabditis elegans
qb_in3:CELT22B11          +     7.00   106.75   1.0e+03   35564  ! AF029040 Caenorhabditis eleg
qb_pr3:AC013256           +     7.00   106.66   1.1e+03   36095  ! AC013256 Homo sapiens chromos
qb_htq25:HSHUCA6          +     7.00   104.28   1.6e+03   44472  ! Z94192 Homo sapiens chromosom
qb_pl3:NCB17C10           -     7.00   102.91   1.7e+03   68484  ! AL355926 Neurospora crassa DN
qb_htq19:AC090603         -     7.00   102.79   1.7e+03   69954  ! AC090603 Homo sapiens chromos
qb_htq5:AC016176          +     7.00   102.36   1.8e+03   75254  ! AC016176 Homo sapiens clone R
qb_htq9:AC024275          -     7.00   102.16   1.9e+03   77866  ! AC024275 Homo sapiens clone R
qb_pl3:NCB24H17           +     7.00   101.79   2.0e+03   82572  ! AL356815 Neurospora crassa DN
qb_htq6:AC018338          -     7.00   101.75   2.0e+03   83548  ! AC018338 Drosophila melanogas
qb_htq20:AL162401         +     7.00   101.44   2.1e+03   88027  ! AL162401 Homo sapiens chromos
qb_htq25:LMFLCHR32_09     +     7.00   100.14   2.4e+03   110000  ! Continuation (10 of 28) of
qb_htq25:LMFLCHR32_10     +     7.00   100.14   2.4e+03   110000  ! Continuation (11 of 28) of
qb_pr1:AC004706           -     7.00    99.36   2.7e+03   125698  ! AC004706 Homo sapiens chrom
qb_pr3:AC011287           -     7.00    99.03   2.8e+03   132943  ! AC011287 Homo sapiens BAC cl
qb_vi2:U93872             -     7.00    99.00   2.8e+03   133641  ! U93872 Kaposi's sarcoma-asso
qb_vi2:KSU75698           -     7.00    98.83   2.9e+03   137508  ! U75698 Kaposi's sarcoma-asso
qb_htq2:AC009381          -     7.00    98.64   2.9e+03   142118  ! AC009381 Drosophila melanoga
qb_pr2:AC005831           -     7.00    98.54   3.0e+03   144515  ! AC005831 Homo sapiens 12p13.
qb_pr9:HSDJ977B1          +     7.00    98.52   3.0e+03   145068  ! AL050318 Human DNA sequence
qb_htq11:AC027529         -     7.00    98.40   3.1e+03   150626  ! AC027529 Homo sapiens chrom
qb_htq7:AC020583          +     7.00    98.40   3.1e+03   150751  ! AC020583 Homo sapiens chrom
qb_htq8:AC022511          +     7.00    98.25   3.1e+03   151865  ! AC022511 Homo sapiens chrom
qb_htq4:AC015688          -     7.00    98.25   3.1e+03   151928  ! AC015688 Homo sapiens chrom
qb_htq10:AC025531         -     7.00    98.16   3.1e+03   154278  ! AC025531 Homo sapiens chr
qb_htq12:AC017984         -     7.00    98.14   3.1e+03   154788  ! AC017984 Homo sapiens chr
qb_htq22:AL390741         -     7.00    97.86   3.3e+03   162334  ! AL390741 Homo sapiens chr
qb_pr4:AC079175           -     7.00    97.83   3.3e+03   163198  ! AC079175 Homo sapiens Xp
qb_htq22:AL513470         -     7.00    97.77   3.3e+03   165072  ! AL513470 Mus musculus chr
qb_htq9:AC024517          -     7.00    97.77   3.3e+03   165079  ! AC024517 Homo sapiens chr

seq_name: qb_patl:AX048148

seq_documentation_block:
LOCUS       AX048148      1894 bp    DNA              PAT       15-DEC-2000
DEFINITION  Sequence 1 from Patent WO0070092.
ACCESSION   AX048148
VERSION     AX048148.1  GI:11876967
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1894)
  AUTHORS   Kaufmann,J. and Xin,H.
  TITLE     Expression of ets-domain proteins in cancer
  JOURNAL   Patent: WO 0070092-A 1 23-NOV 2000;
            Chiron Corporation (US)
FEATURES             Location/Qualifiers
     source          1..1894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             416..1423
                     /note="human JKEIS."
                     /codon_start=1
                     /protein_id="CAC19002.1"
                     /db_xref="GI:11876968"
                     /translation="MGSASQGLSSVQLSNAGLELLEVGKTGLEKAAAGAVGLERRDWS
                     FSFCAIIQGLSAFYISYFDMLYPFGSSWAAKAPGASSEEPEEGPQLFVTDSQAPA
                     GSIDLVPSGLTLEEHSLEQVQSMVVGLVLKDIETACKLLNITADFMLWSFSNVQKWLL
                     WIEHQYELPPMGKAPQELACKELCAMSEPQFRQRSPLQGDVIHAHLDIWKSAAWMKEK
                     TSLGATHYDASISFLSWIPSEVESSPGQLTHEWQLLELLGKPHPNPLVPWRNEEP
                     TFFIDGEASAVAELWCTEENEFAMNYQELSEPTRQYTEECLIPEPLQPSQMLVYQFVHF
                     I"
BASE COUNT       468 a    658 c    571 g    302 t
ORIGIN


alignment_scores:
            Quality:     9.00              Length:       9
              Ratio:    1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_21_29 x AX048148   ..

 Align seq 1/1  to: AX048148  from: 1  to: 1894

        1 AspIleValSerArgIleGlyLeuGln 9
          |||||||||||||||||||||||||||
      476 GACATCGTCTCCCGCATAGGCCTGGGAG 502

seq_name: qb_pr4:AF071538

seq_documentation_block:
LOCUS       AF071538      1894 bp    mRNA             PRI       10-JAN-2000
DEFINITION  Homo sapiens Ets transcription factor EGEF (PGEF) mRNA, complete
            cds.
ACCESSION   AF071538
VERSION     AF071538.1  GI:4007417
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1894)
  AUTHORS   Oettgen,P., Finger,E., Sun,Z., Akbarali,Y., Thamrongsak,U.,
```

```
                Boltax,J., Grall,F., Dube,A., Weiss,A., Brown,L., Quinn,G., Kas,K.,
                Endress,G., Kunsch,C. and Libermann,T.A.
  TITLE         PDEF, a novel prostate epithelium-specific ets transcription
                factor, interacts with the androgen receptor and activates
  JOURNAL       prostate specific antigen gene expression
  MEDLINE       J. Biol. Chem. 275 (2), 1216-1225 (2000)
REFERENCE       20992894
                2  (bases 1 to 1894)
  AUTHORS       Libermann,T.A., Oettgen,P., Kunsch,C. and Finger,E.
  TITLE         Direct Submission
  JOURNAL       Submitted (10-JUN-1998) Medicine, Beth Israel Deaconess Medical
                Center, 330 Brookline Ave., Boston, MA 02215, USA
FEATURES        Location/Qualifiers
     source     1..1894
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="6"
                /map="6p21.3"
     gene       /tissue_type="prostate"
                1..1894
     CDS        /gene="PDEF"
                416..1423
                /gene="PDEF"
                /note="prostate epithelium specific"
                /codon_start=1
                /product="Ets transcription factor PDEF"
                /protein_id="AAC96296.1"
                /db_xref="GI:4007418"
                /translation="MGSASPGLSSVSPSHLLLPPDTVSKTGLEKAAAGAVGLEKKDWS
                PSPPATPEQGLSAKYLSYFDMLYPEDSSWAAKAPGASSKEEPPEEPEQCPVTDSQAPA
                GSLDLVPGGLTLEEHSLEQVQSMVVGEVLKDITACKLLNITADPMDWSPSNVQKWLL
                WTEHQYRLPPMGKAFQELAGKELCAMSEEQFRQRSPLGGDVLHAHLDIWKSAAWMKER
                TSPGATHYCASTSEESWTDSEVDSSCSGQPTHLWQFLKELLLKPHSYGRFIRWLNKEK
                GIFKIEDSAQVARLWGIRKNRPAMNYDKLSRSIRQYYKKGIIRKPDTSQRLVYQFVHP
                I"
BASE COUNT      468 a    653 c    571 g    302 t
ORIGIN


alignment_scores:
            Quality:    9.00
              Ratio:   1.000                Length:       9
Percent Similarity: 100.000   Percent Identity: 100.000    Gaps:       0
alignment_block:
US-09-126-945B-2_COPY_21_29 x AF071538  ..

Align seq 1/1  to: AF071538  from: 1  to: 1894

        1 AspThrValSerArgThrGlyLeuGlu 9
          ||||||||||||||||||||||||||||
      476 GACACGGCTGTCGCGGACAGGCTTGGAG 502

Seq name: qb_pri:AB041549

Seq documentation block:
LOCUS           AB041549        1898 bp    mRNA                 PRI       20-JAN-2000
DEFINITION  Homo sapiens PSE mRNA for prostate ets, complete cds.
ACCESSION   AB041549
VERSION     AB041549.1  GI:6721497
KEYWORDS    prostate ets.
SOURCE      Homo sapiens adult male prostate epithelium prostate cancer
                cell_line:PC-3 cDNA to mRNA.
  ORGANISM  Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1  (sites)
  AUTHORS       Yamada,N., Tamai,Y., Miyamoto,H. and Nozaki,M.
  TITLE         Cloning and expression of the mouse Pse gene encoding a novel Ets
                family member
  JOURNAL       Gene 241 (2), 262-274 (2000)
  MEDLINE       20147504
REFERENCE       2  (bases 1 to 1898)
  AUTHORS       Nozaki,M.
  TITLE         Direct Submission
  JOURNAL       Submitted (25-AUG-1999) to the EMBL/GenBank/DDBJ databases. Masami
                Nozaki, Osaka University, Institute for Microbial Diseases; 3-1,
                Yamadaoka, Suita, Osaka 565-0871, Japan
                (e-mail:mnozaki@biken.osaka-u.ac.jp, Tel:81-6-879-8439,
                Fax:81-6-879-8439)
FEATURES        Location/Qualifiers
     source     1..1898
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /cell_line="PC-3"
                /cell_type="prostate cancer"
                /chromosome="6"
                /dev_stage="adult"
                /map="6p21.2-21.3"
                /sex="male"
                /tissue_type="prostate epithelium"
     gene       /note="vector:pUC18"
                406..1413
     CDS        /gene="PSE"
                406..1413
                /gene="PSE"
                /codon_start=1
                /product="prostate ets"
                /protein_id="BAA89543.1"
                /db_xref="GI:6721498"
                /translation="MGSASPGLSSVSPSHLLLPPDTVSFTGLEKAAAGAVGLEKRDWS
                PSPPATPEQGLSAKYLSYFDMLYPEDSSWAAKAPGASSKEEPPEEPEQCPVTDSQAPA
                GSLDLVPGGLTLEEHSLEQVQSMVVGEVLKEITACKLLNITADPMDWSPSNVQKWLL
                WTEHQYRLPPMGKAFQELAGKELCAMSEEQFRQRSPLGGDVLHAHLDIWKSAAWMKER
                TSPGATHYCASTSEESWTDSEVDSSCSGQPTHLWQFLKELLLKPHSYGRFIRWLNKEK
                GIFKIEDSAQVARLWGIRKNRPAMNYDKLSRSIRQYYKKGIIRKPDTSQRLVYQFVHP
                I"
BASE COUNT      480 a    651 c    567 g    300 t
ORIGIN


alignment_scores:
            Quality:    9.00
              Ratio:   1.000                Length:       9
Percent Similarity: 100.000   Percent Identity: 100.000    Gaps:       0
alignment_block:
US-09-126-945B-2_COPY_21_29 x AB041549  ..

Align seq 1/1  to: AB041549  from: 1  to: 1898

        1 AspThrValSerArgThrGlyLeuGlu 9
          ||||||||||||||||||||||||||||
      466 GACACGGCTGTCGCGGACAGGCTTGGAG 492

Seq name: qb_pat1:AX048159

Seq documentation block:
LOCUS           AX048159        3817 bp    DNA                  PAT       15-DEC-2000
DEFINITION  Sequence 12 from Patent WO0070092.
ACCESSION   AX048159
VERSION     AX048159.1  GI:11876975
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1  (bases 1 to 3817)
  AUTHORS       Kaufmann,J. and Xin,H.
  TITLE         Expression of ets-domain proteins in cancer
  JOURNAL       Patent: WO 0070092-A 12 23-NOV 2000;
                Chiron Corporation (US)
FEATURES        Location/Qualifiers
     source     1..3817
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
```

```
                     /note "GSEF-encoding sequence with promoter (Pins, ZA ZB)"
     promoter        1. .1380
                     /note "P
     CDS             1815. .2822
                     /note="GSEF coding sequence"
                     /codon_start 1
                     /protein_id="CAC19004.1"
                     /db_xref "GI:11876976"
                     /translation "MSSATEP
                     PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAIGASSREELPEEEEQCEVIDSQAFA
                     GSLDLVPGGLTLEEHSLEQVPCMVVCEVLKDTPLACKLLNTLACPWWSPSNVPEWLL
                     WTEHQYRCFPMCKACPCAKLPAMCPSLPLKSCWELTPWLNEEK
                     TSPGATHYCASTSEESWTDSEVDSSCSGQPIHLWQFLKELLLKPHSYCKLTPWLNEEK
                     GIFKTLCSAGVAALWTSLKSAPAMGLSLPAPQVLAPLKSLLSLGPVLPVPL
                     I"
BASE COUNT      719 a   1026 c    970 g    611 t
ORIGIN


alignment_scores:
              Quality:     9.00              Length:       9
                Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_21_29 x AX048159    ..

 Align seq 1/1  to: AX048159  from: 1  to: 3417

        1 AspThrValSerArgThrGlyLeuGlu 9
          |||||||||||||||||||||||||||
     1875 GACACGGTGTCGCGGACAGGCTTGGAG 1901

seq_name: qb_pr6:AL157372

seq_documentation_block:
LOCUS        AL157472    150660 bp    DNA              PRI       29-NOV-2000
DEFINITION   Human DNA sequence from clone RP11-375E1 on chromosome 6, complete
             sequence.
ACCESSION    AL157472
VERSION      AL157472.18  GI:11493236
KEYWORDS     HTG.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 150660)
  AUTHORS    Skuce,C.
  TITLE      Direct Submission
  JOURNAL    Submitted (29-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
             CB10 1SA  UK  E-mail enquiries: humquery@sanger.ac.uk
             requests: clonerequest@sanger.ac.uk
COMMENT      On Nov 30, 2000 this sequence version replaced gi:11344772.
             During sequence assembly data is compared from overlapping clones.
             Where differences are found these are annotated as variations
             together with a note of the overlapping clone name. Note that the
             variation annotation may not be found in the sequence submission
             corresponding to the overlapping clone, as we submit sequences with
             only a small overlap as described above.
             This sequence has been finished according to sequence map criteria
             as follows.  An attempt is made to resolve all sequencing problems,
             such as compressions and repeats, but not necessarily within known
             annotated repeat sequence elements.  Where the sequence is
             ambiguous, there is an annotation using the 'unsure' feature key.
             The following abbreviations are used to associate primary accession
             numbers given in the feature table with their source databases.
             Em:, EMBL; Sw:, SWISSPROT; Ir:, TREMBL; Wp:, WORMPEP; Information
             on the WORMPEP database can be found at
             http://www.sanger.ac.uk/Projects/C_elegans/wormpep
             This sequence
             was generated from part of bacterial clone contigs of human
             chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
             Group.  Further information can be found at
             http://www.sanger.ac.uk/HGP/Chr6
             RP11-375E1 is from the library RPCI-11.2 constructed at the Roswell
             Park Cancer Institute by the group of Pieter de Jong. For further
             details see http://bacpac.chori.org/
             VECTOR: pBACe3.6
             IMPORTANT: This sequence is not the entire insert of c. be
             RP11-375E1 It may be shorter because we sequence overlapping
             sections only once, except for a 100 base overlap.
             The true left end of clone RP11-375E1 is at 1 in this sequence. The
             true left end of clone RP11-302L10 is at 1 in this sequence.
             The true right end of clone RP11-302L10 is at 113944 in this
             sequence.
FEATURES             Location/Qualifiers
     source          1. .150660
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /chromosome "6"
                     /clone "RP11-375E1"
                     /clone_lib "RPCI-11.2"
BASE COUNT     35470 a   38887 c   38740 g   37563 t
ORIGIN


alignment_scores:
              Quality:     9.00              Length:       9
                Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_21_29 x AL157472/rev    ..

 Align seq 1/1  to reverse of AL157472  from: 1  to: 150660

        1 AspThrValSerArgThrGlyLeuGlu 9
          |||||||||||||||||||||||||||
   129680 GACACGGTGTCGCGGACAGGCTTGGAG 129654

seq_name: qb_pr8:HS11B8F

seq_documentation_block:
LOCUS        HS11B8F       288 bp    DNA              PRI       19-OCT-1995
DEFINITION   H.sapiens CpG island DNA genomic MseI fragment, clone 11b8, forward
             read cpg11b8.ftla.
ACCESSION    Z59014
VERSION      Z59014.1  GI:1040927
KEYWORDS     CpG island; genomic MseI fragment.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 288)
  AUTHORS    Bird,A.P., Wilkinson,L. and Mc Klem,G.
  TITLE      Direct Submission
  JOURNAL    Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
             CB10 1RQ, England. E-mail contact: humquery@sanger.ac.uk
REFERENCE    2  (bases 1 to 288)
  AUTHORS    Cross,S.H., Charlton,J.A., Nan,X. and Bird,A.P.
  TITLE      Purification of CpG islands using a methylated DNA binding column
  JOURNAL    Nat. Genet. 6 (3), 236-244 (1994)
  MEDLINE    94282070
COMMENT      Vector: pGEM-5Zf(-)
             Clones are available from the UK MRC Human Genome Mapping Project
             Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See UKL:
             http://www.hgmp.mrc.ac.uk/ for details
             or contact: biohelp@hgmp.mrc.ac.uk.
FEATURES             Location/Qualifiers
     source          1. .288
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /sex "male"
                     /dev_stage="adult"
                     /tissue_type "blood"
                     /clone_lib "CGI-1"
                     /clone "11b8"
```

```
BASE COUNT        60 a      79 c      77 g      64 t       9 others
ORIGIN


alignment scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2 COPY 21_29 x HSI188F   ..

 Align seq 1/1   to: HSI188F   from: 1   to: 288

        8 ValSerArgThrGlyLeuGlu 9
          |||||||||||||||||||||
      200 GTGAGCCGCACCGCTTACAG 220

seq name: qb_in3:HSZ93661

seq documentation block:
LOCUS       HSZ93661       417 bp    DNA              ENV       01 MAY 1997
DEFINITION  Heratemis sp. 28S rRNA gene, D2 variable region.
ACCESSION   Z93661
VERSION     Z93661.1  GI:2065188
KEYWORDS    28S ribosomal RNA; 28S rRNA gene; D2 variable region.
SOURCE      Heratemis sp.
  ORGANISM  Heratemis sp.
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita;
            Ichneumonoidea; Braconidae; Alysiinae; Heratemis.
REFERENCE   1  (bases 1 to 417)
  AUTHORS   Gimeno,C., Belshaw,R.D. and Quicke,D.L.J.
  TITLE     Phylogenetic relationships of the Alysiinae/Opiinae (Hymenoptera:
            Braconidae) and the utility of cytochrome b, 16S and 28S D2 rRNA
  JOURNAL   Insect Mol. Biol. In press
REFERENCE   2  (bases 1 to 417)
  AUTHORS   Belshaw,R.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-1997) Belshaw R.D., Imperial College, Silwood
            Park, Ascot, Berks SL5 7PY, UK
COMMENT     excluding final 20 bp.
FEATURES             Location/Qualifiers
     source          1..417
                     /organism="Heratemis sp."
                     /db_xref="taxon:61232"
     rRNA            <1..>417
                     /gene="28S rRNA"
                     /note="D2 variable region"
                     /product="28S ribosomal RNA"
     gene            1..417
                     /gene="28S rRNA"
BASE COUNT       92 a      84 c      99 g     142 t
ORIGIN


alignment scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2 COPY 21_29 x HSZ93661   ..

 Align seq 1/1   to: HSZ93661   from: 1   to: 417

        2 ThrValSerArgThrGlyLeu 8
          |||||||||||||||||||||
      370 ACGGTATCTAGAACTGGCTTA 390

seq name: qb_st5:CNS06RLJ

seq documentation block:
```

```
LOCUS       CNS06RLJ      1089 bp    DNA              SIS       10-JAN-2001
DEFINITION  3' end of clone AW0AA015E07 of library AW0AA from strain CLIB 89 of
            Yarrowia lipolytica, sequence tagged site.
ACCESSION   AL412109
VERSION     AL412109.1  GI:12182266
KEYWORDS    STS.
SOURCE      Yarrowia lipolytica.
  ORGANISM  Yarrowia lipolytica
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Dipodascaceae; Yarrowia.
REFERENCE   1  (bases 1 to 1089)
  AUTHORS   Casaregola,S., Nouvel-lise,C., Lepingle,A., Bon,E., Feynerol,C.,
            Artiguenave,F., Wincker,P. and Gaillardin,C.
  TITLE     Genomic Exploration of the Hemiascomycetous Yeasts: 17. Yarrowia
            lipolytica
  JOURNAL   FEBS Lett. 487 (1), 95-100 (2000)
  PUBMED    11152892
REFERENCE   2  (bases 1 to 1089)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Nouvel-lise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
            yeast species for molecular evolution studies(1)
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  PUBMED    11152876
REFERENCE   3  (bases 1 to 1089)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            Seqret@genoscope.cns.fr  Web : www.genoscope.cns.fr)
COMMENT     This STS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            cerevisiae, Saccharomyces servazzii, Saccharomyces unisporus,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 4 to
            6 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1..1089
                     /organism="Yarrowia lipolytica"
                     /strain="CLIB 89"
                     /db_xref="taxon:4952"
                     /clone="AW0AA015E07"
                     /clone_lib="AW0AA"
                     /note="end : T7"
     misc_feature    <1..>84
                     /note="similar to Saccharomyces cerevisiae ORF YML041c [
                     hypothetical protein ]"
                     /evidence=not_experimental
BASE COUNT      249 a     298 c     231 g     311 t       1 others
ORIGIN


alignment scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2_COPY 21_29 x CNS06RLJ   ..

 Align seq 1/1   to: CNS06RLJ   from: 1   to: 1089

        2 ThrValSerArgThrGlyLeu 8
          |||||||||||||||||||||
      144 ACGGTATCTAGAACTGGCTTG 164
```

```
seq_name: qb_pl3:PSP276438

seq_documentation_block:
LOCUS       PSP276438    2306 bp    mRNA            PLN       22-MAR-2001
DEFINITION  Piromyces sp. E2 mRNA for beta-glucosidase (bglIA gene).
ACCESSION   AJ276438
VERSION     AJ276438.1  GI:13445201
KEYWORDS    beta-glucosidase; bglIA gene.
SOURCE      Piromyces sp. E2.
  ORGANISM  Piromyces sp. E2
            Eukaryota; Fungi; Chytridiomycota; Neocallimastichales;
            Neocallimasticaceae; Piromyces.
REFERENCE   1  (bases 1 to 2306)
  AUTHORS   Harhangi,H.R., Steenbakkers,P.J.M., Akhmanova,A.S., van der
            Drift,C. and Op den Camp,H.J.M.
  TITLE     A new beta-glucosidase from the anaerobic fungus Piromyces sp. E2,
            belonging to glycosyl hydrolase family 1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2306)
  AUTHORS   Op den Camp,H.J.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Op den Camp H.J.M., Microbiology,
            University of Nijmegen, Toernooiveld 1, Nijmegen, NL-6525 ED,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1..2306
                     /organism="Piromyces sp. E2"
                     /db_xref="taxon:73868"
     gene            1..2306
                     /gene "bglIA"
     CDS             1..1995
                     /gene "bglIA"
                     /EC_number "3.2.1.21"
                     /function "hydrolysis of [illegible]"
                     /codon_start=1
                     /evidence=experimental
                     /product="beta-glucosidase"
                     /protein_id="CAC34952.1"
                     /db_xref="GI:13445202"
                     /translation="MFTSIFTLEAVFAAFFVVFGLSAKLNLVEAPVTLKSG
                     NETPVNVDENTPLDKEEDSGIENSLIENNFELKPEGKLEALEEWFAAIAAVLPVEGAWNE
                     GGEGPSVWDHFTHIYPENVESGDKSNENSINSWNVAPLSYHFLLPSLEMMFIMELNHYF
                     PSMSWSRLPPTSQAEKVESKWNVNFKLVEYVLMVIGCLLPSNLSMATIYHWLEELS
                     MKSMVVGRLPFTRREAKYAEEFCFEHEGDRVKNWITTNEPWVNCVSGYELGEGKAPYRC
                     TGEAPRKLANDTAGLDLLGGCSVEFGFSGSYTKNHEALLANLLDPGRKHPYWC
                     HAKAVNVYRKKFQKKQKGLGGIFLMSLADDWGGDKLPEFLENNSKYANLAAEFPTG
                     WYSDPPMIGDYPESVKQFMNSKDMPVETDPEFKALKKGSVSFIFVWNIYLAHWAIQVKNL
                     DGIYVQPPTDPEFANFENSKFPMWLGNCPTCTWTSPFLSPAASSWNFFYVFIP
                     VGLNWFADRHFSPLIVIIFNCVALGNYFVSRANLEVLFAYTFLGLFFVADVTSF
                     STEDEKNIFGSTLHDFYRDWYKQYLFNLKLAYAVDNVKGYMAWSLLDNFEWENGYETKF
                     YETKFPDMFYDFYNDKNLIVPKPSLEHLGQWYLENVEQN"
     sig_peptide     1..66
                     /gene "bglIA"
     mat_peptide     67..1992
                     /gene "bglIA"
                     /EC_number "3.2.1.21"
                     /product="BGLIA"
     3'UTR           1996..2306
                     /gene "bglIA"
BASE COUNT      806 a    439 c    426 g    635 t
ORIGIN


alignment_scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_21_29 x PSP276438   ..

 Align seq 1/1  to: PSP276438  from: 1  to: 2306

       2 IntValSerArgIhrGlyLeu 8
         ||||||||||||||||||||
     737 ATCGGTGTCAAGAACTGGATTA 757

seq_name: qb_pl1:AF016864

seq_documentation_block:
LOCUS       AF016864     2435 bp    mRNA            PLN       29-DEC-1999
DEFINITION  Orpinomyces sp. PC-2 beta-glucosidase (bglI) mRNA, complete cds.
ACCESSION   AF016864
VERSION     AF016864.1  GI:5639612
KEYWORDS
SOURCE      Orpinomyces sp. PC-2.
  ORGANISM  Orpinomyces sp. PC-2
            Eukaryota; Fungi; Chytridiomycota; Neocallimastichales;
            Neocallimasticaceae; Orpinomyces.
REFERENCE   1  (bases 1 to 2435)
  AUTHORS   Ximenes,E.A., Chen,H., Li,X., Felix,C.R. and Ljungdahl,L.G.
  TITLE     Cloning and sequencing of a beta-glucosidase from the anaerobic
            fungus Orpinomyces sp. strain PC-2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2435)
  AUTHORS   Ximenes,E.A., Chen,H., Li,X., Felix,C.R. and Ljungdahl,L.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUL-1997) Biochemistry and Molecular Biology,
            University of Georgia, Greene Street, Athens, GA 30602-7229, USA
FEATURES             Location/Qualifiers
     source          1..2435
                     /organism="Orpinomyces sp. PC-2"
                     /strain="PC-2"
                     /db_xref="taxon:50059"
                     /note="anaerobic rumen fungus"
     gene            1..2435
                     /gene="bglI"
     CDS             49..2012
                     /gene="bglI"
                     /EC_number="3.2.1.21"
                     /note="hydrolytic enzyme"
                     /codon_start=1
                     /product="beta-glucosidase"
                     /protein_id="AAD45834.1"
                     /db_xref="GI:5639612"
                     /translation="MFTSIVFSALLAVIAAKFIVEVPAVACPALPVVLAPLIAPEDS
                     GVFSCPDGFTLDSTIPYSPDVDMSIVEELPAEFFWGAATAAYQVEGAWDEDGKGESV
                     WDHFIHIYPENVESGDKSKDESINSYNLACDSYHSFLDVEMLKLMNAKYYFPFSISWPR
                     IFPTGQAKKVDKKWVNEKGAEYYDMVINTLKKNGIVPFVTLYHWDLPYALHEKYGGW
                     KNPLMIGDYPESVKQFMNSKDMPVETDPEFKALKKGSVSFIFVWNIYLAHWAIQVKNL
                     DGIYVQPPTDPEFANFENSKFPMWLGNCPTCTWTSPFLSPAASSWNFFYVFIP
                     VGLNWFADRHFSPLIVIIFNCVALGNYFVSRANLEVLFAYTFLGLFFVADVTSF
                     STEDEKNIFGSTLHDFYRDWYKQYLFNLKLAYAVDNVKGYMAWSLLDNFEWENGYETKF
                     YETKFPDMFYDFYNDKNLIVPKPSLEHLGQWYLENVEQN"
BASE COUNT      822 a    424 c    452 g    737 t
ORIGIN


alignment_scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_21_29 x AF016864   ..

 Align seq 1/1  to: AF016864  from: 1  to: 2435

       2 IntValSerArgIhrGlyLeu 8
         ||||||||||||||||||||
     757 ATCGGTGTCAAGAACTGGATTA 777
```

```
seq name: qb pat1:AR049614

seq documentation block:
LOCUS       AR049614      3084 bp    DNA             PAT       29 SEP 1999
DEFINITION  Sequence 3 from patent US 5824545.
ACCESSION   AR049614
VERSION     AR049614.1  GI:6005653
KEYWORDS    .
SOURCE      Unknown
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3084)
  AUTHORS   Kollin,Y., Ripple,P., Carrier,V., Bulawa,C. and Winter,F.
  TITLE     Identification of eukaryotic growth related genes and promoter
            isolation vector and method of use
  JOURNAL   Patent: US 5824545-A 3 20-OCT 1998;
FEATURES             Location/Qualifiers
     source          1..3084
                     /organism="unknown"
BASE COUNT       944 a    612 c    550 g    966 t      i others
ORIGIN


alignment scores:
            Quality:     7.00          Length:       7
              Ratio:    1.000            Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_21_29 x AR049614/rev    ..

 Align_seq 1/1  to reverse of: AR049614  from: 1  to: 3084

       2 ThrValSerArgThrGlyLeu 8
         ||||||||||||||||||||||
    1258 ACTGTATCCCGGACCCGGTTTA 1238

seq name: qb vil:AF075255

seq documentation block:
LOCUS       AF075255     11411 bp    RNA             VRL       10 FEB 1999
DEFINITION  Venezuelan equine encephalitis virus strain 71D 1252, complete
            genome.
ACCESSION   AF075255
VERSION     AF075255.1  GI:4262411
KEYWORDS    .
SOURCE      Venezuelan equine encephalitis virus.
  ORGANISM  Venezuelan equine encephalitis virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
            Alphavirus.
REFERENCE   1  (bases 1 to 11411)
  AUTHORS   Kinney,R.M., Pfeffer,M., Tsuchiya,K.K., Chanel,C.J. and Weaver,S.C.
  TITLE     Nucleotide sequences of the 26S mRNAs of the viruses defining the
            Venezuelan equine encephalitis antigenic complex
  JOURNAL   Am. J. Trop. Med. Hyg. 59 (6), 952-964 (1998)
  MEDLINE   99101297
REFERENCE   2  (bases 1 to 11411)
  AUTHORS   Kinney,R.M., Pfeffer,M. and Weissner,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUN-1998) Arbovirus Diseases Branch, Centers for
            Disease Control and Prevention, P.O. Box 2087, Fort Collins, CO
            80522, USA
FEATURES             Location/Qualifiers
     source          1..11411
                     /organism="Venezuelan equine encephalitis virus"
                     /strain="71D-1252"
                     /db_xref="taxon:11036"
                     /country="Peru"
                     /note="isolated in 1971
                     sequencing of RT/PCR cDNA amplicons from plaque purified
                     virus"
     5'UTR           1..43
     CDS             44..7487
                     /note="possible incorporation of arginine, cysteine, or
                     tryptophan at read through UGA codon in nsP4"
                     /codon_start=1
                     /transl_except=(pos:5543..5545,aa:OTHER)
                     /product="nonstructural polyprotein"
                     /protein_id="AAD14558.1"
                     /db_xref="GI:4262412"
                     /translation="MEKVHVPFEPEQSRTCQEIRASVPHDAIQPSLRVAVKTVRTGDWVNCGSIS
                     GLASDSVTYECPMLAENEPDIDYCCRKTCKSTEVNSEYFCKPKKITPTGKRCLALNLATKDFEEY
                     NCWNTSCFTTGKACGAPENYKSDLVDVPEIRTIVAADIHSFWDSCFSQYCHDWGTLKYEWDNEP
                     LGFYHQANKGVKVAYWIGFDTTPFMFKNLAGAYPTYSTNWADITVLTARNCSLYCPEEFLNPVD
                     EGSKCGLSVLRKKYLKPSNNVLFSVGSTIYHEKKDLLKGWHLPSVFHLRGKQNYTCRC
                     ETIVSCDGYVVKRIAISPGLYGKPSGYAATMHREGFLICKVTDTLNGERVSFPVCTYV
                     PATLCDQMTGILATDVSADDAQKLLVGLNQRIVVNGRTQRNTNTMKNYLLPVVAQAFA
                     RWAKEYKEDQEDEKPLGLRDKQLVMGVCWAFKKHRTSVYKRPDTQTVIKVPSDFHSF
                     VLPRLGSCSLETGLKSKTKKLLEDKREAPPLTSPEDVSAAKEAAAEAKETKEAFELRV
                     AETRTVPTVEEPTLEAEVDLMLQEAGAGSVETPRGLIKVTSYSGEDKIGSYAVLSPQA
                     VLKSEKLSCIHPLAEEAIVITHSGRKGRYAVEPYHGKVVVPEGHAIPVADFQALSESA
                     TIVYNEPEIVNRYLHHIAVHGGALNTDEHYYRVVRPGAYTLYGTPPQCVPEIQV
                     SGIALLGLVEPETFHHLEAYESLKTRPAAPSQVPTLGVYGVPGSGKSGIIKSAVTKRDL
                     VVSAKKENCAEIMRDVKKLRGLDIVARTVDSVLLNGCKHPVDTLYIDEAFACHAGTLR
                     ALIAIIRPKKAVLCGDPKQCGFFNMMCLKVHFNHEICTQVFHKSISRRCTKSVTSVVS
                     TLFYDKLMKTNPEPEIEIVLIDGQSYPLCDNPSLIFPAWKEDKISDQCNEIMAAA
                     ASQGLLRKCIYAVPYEKVNEHPLYAPNSEHVNVLLTRTEDRLVWKTLAGDPWIKQLTA
                     YSGDFSATLEEWQQEHDAIMRHILERFPDAFVFQNKANVCWAKALVPVLETAGIDLTA
                     EQWNTVAEFKGDKAHSAEIVLNQLCVKFFGLDLDSGLFSAPTVPLSIRNNHWDNSPSP
                     NMYGLNKEVVRQLAKRYPQLPKAVATGKVIDMNTGTFHKGYLFSVNGLVTVMSKGKLPHALV
                     THKTEQPFESDFSAFASKLKGKTVLVTGDKMHTTGKTVDWLSESPFALFKHRLLDGIPS
                     ELPKYDIVFVNIRTPYKYHHYQQCEDHAIKLSMLAKKACTHLNICGGLCVSVGYGYADK
                     AGSAVVGAVRQLKELSEVCKFKVSLLSGTVLVLEGVEYFTLSRTLTNIYT
                     GSKLHEAGCAPSYHVVRGDIANAEEGVIVNAANSKGQPGKGVCGALYKKFPENFDLQP
                     IEVGKSRLVKGGVKHIIHAVGPNFSKVSEVEGDKQLADAYESVAKILNDNHYRSVAIP
                     LLSTGIFAGSKDRLMQSLNHLLTALDTTDADVAIYCRDKKWEITLKEVIARREAVEEI
                     CISEDASQVEPDAELVRVHPKSSLAGRKGFSTTDGKTFSYLEGTKFHQAAKDVAEINA
                     MWPNAAEANEQICLYIIGESMSSIRSKCPVEESEASTPPNTLPCMCTIHSMTPERVHRL
                     KAARPEQITVCSSFPLPKYRITGVQKIQCAHPTLFLFKVPEYIHPRKYLFNPERSEIQ
                     ETLPDVQESIPEPVTSVTDSLSVVSESTVASTIPNETVVEAEVHSELMREFSSNWSIPQ
                     ASDFDIDSVSVQGFIDIESSAMQEDIRSLASSNSSGIKPTPAPRTVTIQPEPVPKPRRR
                     RRPESSASSTTPLVRTSCATSRSSLVSNPPGERRVITREEFEAFVAQQQXRYDAGAYI
                     FSSDTGQGHLQQKSVRQTVLSEIVLEKTELEIAYAPRIDLEKEEALFKKKLQVNPTPAN
                     RSRYQSRKVENMKAVTFRRTLSGLGHYLRSEGKVECYRTLYPVPLYSASVDRAFSSER
                     VAIEACNRPVLKENEPPTVASYCTTPEYDAYMDMVDCASCCLDIASFCPAKLRSFPKRHA
                     YLEPTIRSAVPSAIQNTLQNVLAAATKRNCNVTQMRELPVLDSAAFNVECFKKYACNN
                     EYWETFKENPIRLTDENVTNYITKLKGPKAAALFAKTHNLSMLQDIPMDRFVMDLKRD
                     VKVTPGTKHTEERPKVQVIQAADPLATAYLCGIHRELVRRLNAVLLPNVHTLFDMSAE
                     DFDAIIAEHFQPGDWVLETDIASFDKSEDDAMALTALMILEDLGVDQELLTLIEAAFG
                     EISSIHLPTKTKFKFGAMMKSGMFLTLFINTVINIVIASRVLRERLTNSPCAAFIGDD
                     NIVKGVKSDKLMADRCATWLNMEVKIIDAVVGEKAPYFCGGFILCDTVTGTACRVADP
                     LKRLEFSCKPLATDEFDEQRRPAIYEESTLRWNRVGIFSELCKAVESKYETTGTSVII
                     AAMTTLARNVAAFKHLRGNPVTLYG"
mat_peptide     44..1648
                     /note="guanylyltransferase"
                     /product="methyltransferase nsP1"
mat_peptide     1649..4030
                     /note="RNA helicase"
                     /product="nonstructural proteinase nsP2"
mat_peptide     4031..5653
                     /product="nonstructural protein nsP3"
mat_peptide     5654..7384
                     /product="RNA polymerase nsP4"
mRNA            7393..11411
                     /note="26S mRNA; structural polyprotein is translated from
                     an intracellular subgenomic 26S mRNA"
                     /product="structural polyprotein"
mat_peptide     7424..8247
                     /note="capsid"
                     /product="nucleocapsid protein"
CDS             7424..11187
                     /codon_start=1
                     /product="structural polyprotein"
                     /protein_id="AAD14559.1"
                     /db_xref="GI:4262413"
                     /translation="MFPYQQTMYFPFQPMPLFKNPYAFFERPWFPKTDFTLAMKVQELAR
```

```
                     SMANLTFKQRRSIPFEGDPAKKPRKEPSQQAIQAC
                     NPKKKVNKKPGKPQPMVMKLESQKTFPMLCLNGYAVVV
                     TLASLKKAASKYDLEYADVKQNMKAQLKKYTPFNQCYYVGWHGTAVQYNCKLTVV
                     GVCAKGDSGRPTLDNQGKVVAIVLGGVNEGSRTALSVVMWNEKGVTVKYTPENCEQWS
                     LVTTMCLLANVTFPCSAPPMCYDRKPAETLAIISANTSNPGVSYDLKAVLFKPKQKF
                     SILLEPSYKLIRDYIAPCVPCAVDSQSLSGAAKAVSSLVWKHLSKYRKPSSDYCLE
                     GNLKGPKMPYDVGQKLKPTPLDRGWSFDLKDLPVGEPGDCGLTMREFF
                     DSIIHSCSVPYEVKFTPVGKELYNEPREFLIEVDNKVVHKAQLKALAYVLMHEKPLSEV
                     DSALLSMSQSAVKVTPPAQQSVSVSKPGKAIKPPPKRVIRTGLRTPTAPCEAYEVLKN
                     DKWVPNSPEVLPFCVKPELKSELRVPYYLSSAPCTVPLAHFQVDIFCTEPVFVKLKPHPN
                     PIYLTTRHLNGEPNYTHELISEPETRNFTVTPQGWEYVWGNHPPQRYWAQETAPGDPH
                     GLPHEVITHYYFKRYPMSTILGLSICAAVVTVAVAASTWLLCKSRACTLTPYRLTPNAR
                     LPVCLAVLCCARAAPAFTLWESLPHLWNFNQMFWSQLLIPLAALIVVTRLLKCLCCV
                     VPFLVLAGAASVGAYEHATTMPNQAGISYNTVVNRSAGYAPVSTTLTKVKLGTGPG
                     EYTTCHYKTGLDSPSIKCCGAQECPKVVRPDPQCKVFTGVYPFMWGGAYCFCDAENTQ
                     VSKAYVTKSSMKSTHHALAYFAHTAPVLAFLNTTKSLNLAVVYVNHETLSFNDVF
                     LVAGPLSTAWTPFDNKVVVYAGEIYNYDFPEYGAGHAGAFDGDLQARFTTSNDLYASTN
                     LILQPPNAGTIHVPYTQAPSGFKYWLKERGASLQHTAPFGCQIATNPVRAENCAVGS
                     PLAFDIPDALFTRVSETPTLSDAGCSLNFCVYSSLPVDTIAVKYSASKAGKCAVHSPS
                     GTATLKFSIVDVPVQESTTHFSTANIHEFKELTCLKCDGTPFPCHPERTGLCDPFCHEYKHV
                     HHAQTFTAAVSKTAWTWLSSLLGGSAATITIGLVLATVAMYVLINQKHN"
     mat_peptide     8248..8424
                     /product="polypeptide cleavage product E3"
     mat_peptide     8425..9690
                     /note="main neutralization epitopes: E2"
                     /product="envelope glycoprotein 2"
     mat_peptide     9691..9858
                     /product="6K polypeptide cleavage product"
     mat_peptide     9859..11181
                     /note "contains fusion domain: E1"
                     /product "envelope glycoprotein 1"
     3'UTR           11185..11411
BASE COUNT      3113 a   2825 c   2928 g   2345 t
ORIGIN


alignment_scores
            Quality:     7.00              Length:       7
              Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block
 US-09-126-945B-2_COPY_21_29 x AF075259_1

 Align seq 1/1  to reverse of  AF075259  from 1  to 11411

        2 ThrValSerArgThrGlyLeu 8
          |||||||||||||||||||||
      944 ACGGTTTCCCGTACAGGCCTG 924

seq_name: db_vil:AF075259

seq_documentation_block:
LOCUS       AF075259     11385 bp    RNA             VRL       10-FEB-1999
DEFINITION  Venezuelan equine encephalitis virus strain Cabassou CbAr 508,
            complete genome.
ACCESSION   AF075259
VERSION     AF075259.1  GI:4262423
KEYWORDS    .
SOURCE      Venezuelan equine encephalitis virus
  ORGANISM  Venezuelan equine encephalitis virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
            Alphavirus.
REFERENCE   1  (bases 1 to 11385)
  AUTHORS   Kinney,R.M., Pfeffer,M., Tsuchiya,K.R., Chanq,G.J. and Koehrig,J.T.
  TITLE     Nucleotide sequences of the 26S mRNAs of the viruses defining the
            Venezuelan equine encephalitis antigenic complex
  JOURNAL   Am. J. Trop. Med. Hyg. 59 (6), 952-964 (1998)
  MEDLINE   99101297
REFERENCE   2  (bases 1 to 11385)
  AUTHORS   Kinney,R.M., Pfeffer,M. and Meissler,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUN-1998) Arbovirus Diseases Branch, Centers for
            Disease Control and Prevention, P.O. Box 2087, Fort Collins, CO
            80522, USA
FEATURES             Location/Qualifiers
     source          1..11385
                     /organism="Venezuelan equine encephalitis virus"
                     /strain="Cabassou CbAr 508"
                     /db_xref="taxon:11036"
                     /country="Guyana"
                     /note "isolated in 1968
                     sequencing of RT/PCR cDNA amplicons from plaque purified
                     virus"
     5'UTR           1..48
     CDS             49..7404
                     /note "possible incorporation of arginine, cysteine, or
                     tryptophan at read-through UGA codon in nsP3"
                     /codon_start=1
                     /transl_except=(pos:5560..5562,aa:OTHER)
                     /product="nonstructural polyprotein"
                     /protein_id="AAD14566.1"
                     /db_xref="GI:4262424"
                     /translation="MEKVHVDVEEDSPFSFLQSLRAGAFGRSDLQVETDKFICENPGVEAKQFTKLVTVNQLPHANALFES
                     SLAAPFQFEDVPFHSTYAHSDTLDQMFLGLYGSAVASFAYMYSLVNKDKLPVSGNMIVSFAQVCTKVTPSRVYFNPYLRE
                     NCKDFLDEGLFQKIADLAAVMGFPSLERESTCLHDDESCRYEGQIAVYQDVYAVDGPT
                     SLYHQANKGVRVAYWIGFDTTPFMFKNLAGAYPSYSTNWADTEVIAARNIGLCSSDVM
                     ERSRRGLSILRKKYLKPSNNILFSVGSTIYHEKRDLLRSWHLPSVFHLRGKQNYTCRC
                     ETIVSCDGYVVKRIAISPGLYGKPSGYAATMHREGFLCCKVTDTLNGERVSFPVCTYV
                     PATLCDQMTGILATDVSADDAQKLLVGLNQRIVVNGRTQRNTNTMKNYLLPVVAQAFA
                     RWAKEYKEDQEDEKPLGLRDRQLVMGCCWAFRKHRITSVYKRPDTQTIVKVPSTDFHSF
                     VLPRVGSSTLEIGLRDKIKNLLEKPKKFVPLVTIDEDIGEAKNAAAFAKEIKEAEELKA
                     ALPFLLIAIVLLFITLAIVSLMKLQEAGAGSVEIPPCTLKVISYSGEEKDGSYATLSPQA
                     VLKSEKLACTHALAEQVIVITHSGRKGKYAVEPYHGKVIVPEGTAIPVEDFQALSESA
                     TIVYNEREFVNRYLHHIAVHGGALNTDEEYYKTVKPSDHDGKYLYDIDNKQCVKKELV
                     SGLGMTGALLHPFIHLFAYLSLNTPAARDQVETLGGVYVLPRASLNTRNAVIKKEDL
                     VVSAKEENCAELTPVKRAGLDVNAPTVDSVLLNGCKPVHVETLYIPDEAFACHAGTLK
                     ALIATVKPKKAVLGGFEKQFGFENMMCLKVHFNHEIDTQVFHKSTPKFCTKSVISVVS
                     TLFYDKKMRTTNQEFQTETIDTDGSTEPKKEFLTLFCFPMVWPQTGYPKGNFIMTAA
                     ASQCLTKEGVVAVPEKAVNENPLYALTNSGHVNVLLTPTEDPTVWETLAGDSWFETLIAR
                     YPGEFSATLEEWQAEHDAIMKHIMEEPNFAEVFQNKANVCWAKALVPVLKTAGIDLTA
                     PQWNTVDNFVTIPAARSAFIVLNQLCVPTPCLLDPSTIPAAHVPLLPYNNHWLNSPSP
                     NMYGLNMPVVRQLSRRYPQLPRAVIIGRALDMNSGALRAYIPRTNLVPMNKRLPHALV
                     THHSEHPPSDSFIEVSKLKGRIVLVVGDKMIVTGKTVDWLSESPEAIFTFRLDIGIPA
                     FLPEVDFVEVNVPETPYFVHHYQGCPDHATPLSMLTKKACLHLNPGTPCVSVGVPAPR
                     AAGGLLAVAEPLFLAEVGDSIFKISLEETEVLFVPLGYDPKFRTHNFYKLSSTLLNTYS
                     GSNFHEAGCAPSYHVVRGDIATAEDGVIVNAANSRGQPGGGVCGALYKKFPESFDLQP
                     IEVGKARLVKGAAKHIIHAVGPNFNKVSEVEGDKQLAEAYESVAKLVMINNYESVAIP
                     LLSTGIFAGNKDRLMQSLNHLLTAMDTTDADVAIYCRDKKWEMTLKEAVARREAVEEI
                     CISEDESLTEPIVECEVRVPPKSSLAGRQGYSTIDGFTFKSCVEQTAKLTDVAKFNMT
                     MWPNATLANEVCTCLYTLGESMSSIRSKCPVEESFASTPPNTLPCLCTIHAMTPERVQRL
                     KASPPEQITVCSSFPLPKYKIEGVQKIQCSCPTIFSPEVPEYTHFKKYLPRFLTEVVE
                     KAGTPAPASNVEAATFENPFETSVVDSVVSSSESTAETNLIGPNATVVFAIVDHADLMRF
                     SSIWSIPHASDFDADSISLALSEFSSIDAKSVVSSSSSIAFTKFILAFPKTIFDQAPVP
                     KPRKTRSSSASLSPTHARPQVSGSRSSLASNPPGVNKVIIREEFEAFVAQQQKRYEA
                     GAYIFSSDTGQGHLQQKSVRQTVLSEVVLETLLTLDVYAGSFDLNKEEIEEQTKATGLNP
                     TGANRSKYQSRKVENMSAITIMEMLEGLGHYLKTDNKVEGYRTLYYPVPLEYSSSVNRSF
                     SCPEVAVEACNVVLKENPPIVASYGIIPEYDAYLDMVDGASCCLNTASFCPAKLRSFP
                     KKHSYLEPTIRSAVPSAIQNTLQNVLAAAIKKNCNVTQMPELPVLDSAAFNVECFKKY
                     ACNNEYWDTYRDNPIRLTEENVTNYITKLKGPKAAALFAKTHNLCMLQDIPMDRFVMD
                     LKRDVKVTPGTKHTEERPKVQVIQAADPLATAYLCGIHRELVRRLNAVLLPNVHTLFD
                     MSAEDFDAIIAEHFQPGDQVLETDIASFDKSEDDAMALTALMILEDLGVDQELLTLIE
                     AAFGEISSIHLPTKTKFKFGAMMKSGMFLTLFVNTLLNIVIASRVLRERLTNSPCAAF
                     IGDDNIVKGVKSDKLMADRCATWLNMEVKIIDAVVGEKAPYFCGGFILCDSVTGTACR
                     VADPLKRLFKLGKPLAAGDEHDDQRRALHEESTRWNRAGIVSELCKAVESRYETVGT
                     SIIIAAMAILARNVAIFKHLRGNPIHLYG"
     mat_peptide     49..1653
                     /note "guanyltransferase"
                     /product="methyltransferase nsP1"
     mat_peptide     1654..4035
                     /note "RNA helicase"
                     /product="nonstructural proteinase nsP2"
     mat_peptide     4036..5580
                     /product "nonstructural protein nsP3"
     mat_peptide     5581..7401
                     /product "RNA polymerase nsP4"
```

```
     mRNA            7410. .11385
                     /note "26S mRNA; structural polyprotein is translated from
                     an intracellular subgenomic 26S mRNA"
                     /product "structural polyprotein"
     mat_peptide     7440. .8264
                     /note "capsid"
                     /product "nucleocapsid protein"
     CDS             7440. .11204
                     /codon_start 1
                     /product "structural polyprotein"
                     /protein_id "AAD14567.1"
                     /db_xref "GI:4262325"
                     /translation "MEPVQPSMYPMQPAPYRPYPAPRRPWYPRTLPFLAIQVQPLAPC
                     MANLTFKQRKESTPEGPPAEKKKKPQQAATPKNAQEKNGKCKPKPKPCAVQPFNQP
                     ASKKKPNKEPCPKPQPMVMKLESPKTPIMEGKINCYAVVGGKGFKGMVPCKDNE
                     TLAALETKEAIPVPLAYAEVPQLMRAGIIEYIHLKPQCYVSWRHCAVQYLNCGETLVCR
                     GVGAKGDSGRPTLDNQGKVVAIVLGGVNEGSKTALSVVMWTEKGVTVKYTPENCEQWS
                     LVTAVCLLANVTFPCAQPFCYLEELTELAMLCLNTCMFGYLVELGCVCECPCPQLH
                     EVTYKPKKQTAWAKCPKAGLRPLCVEEVAGSALLVEDGVNKPAARERLSEDGHFSV
                     GGVKSRTMRYNIQQGNIEETLPLHEVSLHTSRPCHIIDGHGYFFLLARCPEGDSLTMEFFK
                     DTVTHSCSVPYKVKFIPVGKELYTHPPEHGIDHPCKVYAHDAQKRGAYVEMHLPGSEV
                     DSTLLSMSCCAVQVNPPATCNVIVEPNPCPQTSPTVCPVKFTNQTSFNPTFAYTKQS
                     DKWVENGDKLFKASCPLFEGFPHVPLTLSLAPCTVPLAFLEVVSLCPPAVLPFSHCNN
                     PTYLTTRHLGGEPQQYTHELISEPVVKNFSTTEKGWEFVWGNHPPQRFWAQETAPGNPH
                     GMPHEIVTHYYYRYPMSTVVGLSICAAIVLFSTAASICLLCKSRVSCLTPYRLTPNAR
                     LPICLALLCCARPTRAETTWETLGHEWNNNQQMFWIQLLIPLAALIVITRTLKCVCTP
                     VPFLVLAGAAGAGAYEHATTMPSQVGIPFNTTVNRAGYAPLATSIIPTKIQLTPTLNL
                     EYITCHYKTGIDSPAVKCCGSQECSEVTPPDHFTGVYPEMWDCAYCPQVTFNSQ
                     VSKAYVMKSDNWAAIHAQAYKAIRVAAAQATLNTTVAFGGLFTIVYVNGLFYVNFNGIR
                     ETAGPLSTAWTFPEDRKIVQYAEEVYNYDFPEYGAGIAGAFGDLQARTLSTDLYANTN
                     LVLQRPEPSSIVHVPYPQARSCFEGWFKGKPIPSIEYTAIGTCEQKVNPIRAENCAVGSI
                     PLSFDIPDALFTRVSETPTLSNAEFTFLNLCVYSSTGFKIASEVKEVSATKAGEGAVHTPS
                     GTATVREATVDLVEQGSETVHFSTANIHPEFRLQICTTYVTCKGDCHPPKDHIVTHPQ
                     YHAQTFTAAVSKTAWTWLTSLIGGSAVIVVIGLILATIVAIYVLIQQRRY"
     mat_peptide     8265. .8441
                     /product "polyprotein cleavage product E3"
     mat_peptide     8442. .9707
                     /note "main neutralization epitopes; E2"
                     /product "envelope glycoprotein 2"
     mat_peptide     9708. .9875
                     /product "6K polyprotein cleavage product"
     mat_peptide     9876. .11201
                     /note "contains fusion domain E1"
                     /product "envelope glycoprotein 1"
     3'UTR           11202. .11385
BASE COUNT      3244 a   2795 c   2814 g   2532 t
ORIGIN


alignment scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945b-2_copy_21_29 x AF156545/7410-

 Align Seq 1/1   1  ACGGTGTCCCGTACAGGCCTG 929

       2 ThrValSerArgThrGlyLeu 8
         |||||||||||||||||||||
     949 ACGGTGTCCCGTACAGGCCTG 929

seq name: gb_pr5:AF156545

seq documentation block:
LOCUS       AF156545     30461 bp    DNA             PRI       28-SEP-1999
DEFINITION  Homo sapiens alpha 1 antitrypsin and corticosteroid binding
            globulin intergenic region sequence.
ACCESSION   AF156545
VERSION     AF156545.1  GI:5712740
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 30461)
  AUTHORS   Rollini,P., Namciu,S.J., Marsden,M.D. and Fournier,R.E.
  TITLE     Identification and characterization of nuclear matrix attachment
            regions in the human serpin gene cluster at 14q32.1
  JOURNAL   Nucleic Acids Res. 27 (19), 3779-3791 (1999)
  MEDLINE   99412420
REFERENCE   2  (bases 1 to 30461)
  AUTHORS   Namciu,S.J., Marsden,M.D., Rollini,P. and Fournier,R.E.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUN-1999) Division of Basic Sciences, Fred Hutchinson
            Cancer Research Center, 1100 Fairview Ave N., Seattle, WA
            98109-1024, USA
FEATURES             Location/Qualifiers
     source          1. .30461
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /chromosome "14"
                     /map "14q32.1"
     misc_feature    1. .30461
                     /note "alpha 1 antitrypsin and corticosteroid binding
                     globulin intergenic region"
     repeat_region   247. .572
                     /rpt_type tandem
                     /rpt_unit tq
     repeat_region   656. .710
                     /rpt_type tandem
                     /rpt_unit tq
     repeat_region   774. .957
                     /rpt_type tandem
                     /rpt_unit tq
     repeat_region   1745. .1845
                     /rpt_family "SINE/MIR"
                     /rpt_type dispersed
     misc_feature    2613. .6334
                     /note "ATR nuclear matrix attachment region"
     repeat_region   2847. .4773
                     /rpt_family "LINE/L1"
                     /rpt_type dispersed
     repeat_region   3774. .4049
                     /rpt_family "SINE/Alu"
                     /rpt_type dispersed
     repeat_region   4050. .4109
                     /rpt_family "LINE/L1"
                     /rpt_type dispersed
     repeat_region   4141. .4454
                     /rpt_family "LTR/MER4 group"
                     /rpt_type dispersed
     repeat_region   4459. .4706
                     /rpt_family "LINE/L1"
                     /rpt_type dispersed
     repeat_region   4709. .4825
                     /rpt_family "LTR/MaLR"
                     /rpt_type dispersed
     repeat_region   4828. .5049
                     /rpt_family "LINE/L1"
                     /rpt_type dispersed
     repeat_region   5050. .5555
                     /rpt_family "LINE/L1"
                     /rpt_type dispersed
     repeat_region   5801. .6122
                     /rpt_family "LINE/L1"
                     /rpt_type dispersed
     repeat_region   6441. .6976
                     /rpt_family "LINE/L2"
                     /rpt_type dispersed
     repeat_region   7058. .7080
                     /note "low complexity AT rich"
                     /rpt_type dispersed
     repeat_region   7183. .7222
                     /rpt_family "DNA/Mariner"
```

```
                     /rpt_type dispersed
     repeat_region   7378. .7526
                     /rpt_family "MER103"
                     /rpt_type-dispersed
     repeat_region   8445. .8515
                     /rpt_type tandem
                     /rpt_unit ttca
     repeat_region   8625. .8932
                     /rpt_family "SINE/Alu"
                     /rpt_type dispersed
     repeat_region   9041. .9196
                     /rpt_family "DNA/MER1 type"
                     /rpt_type dispersed
     repeat_region   9374. .9797
                     /rpt_family "LTR/MaLR"
                     /rpt_type dispersed
     repeat_region   10189. .10470
                     /rpt_family "SINE/MIR"
                     /rpt_type dispersed
     repeat_region   10720. .10748
                     /rpt_family "DNA/MER1 type"
                     /rpt_type dispersed
     repeat_region   10749. .11112
                     /rpt_family "LTR/MaLR"
                     /rpt_type dispersed
     repeat_region   11113. .11279
                     /rpt_family="DNA MER1-type"
                     /rpt_type dispersed
     repeat_region   11902. .11927
                     /rpt_type tandem
                     /rpt_unit tg
     repeat_region   12506. .12603
                     /rpt_family="LINE/L2"
                     /rpt_type dispersed
     repeat_region   12797. .13339
                     /rpt_family="LTR/retroviral"
                     /rpt_type dispersed
     repeat_region   13480. .13520
                     /note "low complexity GA rich"
                     /rpt_type dispersed
     repeat_region   14077. .14169
                     /rpt_family="SINE/MIR"
                     /rpt_type dispersed
     repeat_region   14236. .14634
                     /rpt_family="LTR/retroviral"
                     /rpt_type dispersed
     repeat_region   15214. .15520
                     /rpt_family="SINE/Alu"
                     /rpt_type-dispersed
     repeat_region   16053. .16167
                     /rpt_family-"MER103"
                     /rpt_type dispersed
     repeat_region   16440. .16537
                     /rpt_family-"DNA/MER1type"
                     /rpt_type dispersed
     repeat_region   16734. .16808
                     /rpt_family-"DNA/MER1 type"
                     /rpt_type dispersed
     repeat_region   17146. .17425
                     /rpt_family "DNA/MER1 type"
                     /rpt_type dispersed
     repeat_region   17812. .18194
                     /rpt_family="LINE/L1"
                     /rpt_type dispersed
     repeat_region   18481. .18605
                     /rpt_family "SINE/MIR"
                     /rpt_type dispersed
     repeat_region   19156. .19249
                     /rpt_family="SINE/MIR"
                     /rpt_type-dispersed
     repeat_region   20397. .20465
                     /rpt_family "LINE/L2"
                     /rpt_type dispersed
     repeat_region   20758. .21591
                     /rpt_family "LINE/L2"
                     /rpt_type dispersed
     repeat_region   21592. .21636
                     /rpt_type tandem
                     /rpt_unit taaa
     repeat_region   21637. .21792
                     /rpt_family="LINE/L2"
                     /rpt_type dispersed
     repeat_region   21793. .21850
                     /rpt_family="MER81"
                     /rpt_type dispersed
     repeat_region   21851. .22706
                     /rpt_family "LINE/L2"
                     /rpt_type dispersed
     misc_feature    22043. .>30461
                     /note "CHS nuclear matrix attachment region."
     repeat_region   22810. .23195
                     /rpt_family "LINE/L1"
                     /rpt_type dispersed
     repeat_region   23156. .23340
                     /rpt_family "SINE/Alu"
                     /rpt_type dispersed
     repeat_region   23341. .23648
                     /rpt_family "SINE/Alu"
                     /rpt_type dispersed
     repeat_region   23649. .24493
                     /rpt_family "LINE/L1"
                     /rpt_type dispersed
     repeat_region   24588. .24930
                     /rpt_family "LTR/MaLR"
                     /rpt_type dispersed
     repeat_region   24932. .26046
                     /rpt_family "LINE/L1"
                     /rpt_type dispersed
     repeat_region   26054. .26528
                     /rpt_family "LINE/L1"
                     /rpt_type dispersed
     repeat_region   26617. .26736
                     /rpt_family "LINE/L1"
                     /rpt_type dispersed
     repeat_region   26767. .27145
                     /rpt_family "DNA/MER2-type"
                     /rpt_type dispersed
     repeat_region   27150. .27498
                     /rpt_family "LINE/L1"
                     /rpt_type dispersed

alignment_scores:
              Quality:    7.00              Length:       7
                Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_21_29 x AF156545   ..

 Align seq 1/1  to  AF156545  from: 1  to: 30461

        3 ValSerArgThrGlyLeuGln 9
          ||||  ||||||||||||||||
    14084 GTCAGCAGAATGGCCTGGAA 14104

seq_name: gb_pat1:AR065852

seq_documentation_block:
LOCUS       AR065852     22267 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 20 from patent US 5849564.
ACCESSION   AR065852
VERSION     AR065852.1  GI:5996068
KEYWORDS
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
```

```
REFERENCE   1  (bases 1 to 32207)
  AUTHORS   Chang,Y., Bohenzky,R.A., Russo,J.J., Edelman,I.S. and Moore,P.S.
  TITLE     Polypeptides from Kaposi's sarcoma associated herpesvirus, DNA
            encoding same and uses thereof
  JOURNAL   Patent: US 5849564 A 20 15 DEC 1998;
FEATURES             Location/Qualifiers
     source          1..32207
                     /organism "unknown"
BASE COUNT     7229 a    9156 c    8714 g    7109 t
ORIGIN


Alignment scores:
              Quality:     7.00              Length:       7
                Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

Alignment block:
 US-09-126-945B-2_COPY_21_29 x AR065852/rev  ..

 Align seq 1/1  to reverse of: AR065852  from: 1  to: 32207

        2 ThrValSerArgThrGlyLeu 8
          |||||||||||||||||||||
    1649 ACTGTCTCACGGACCGGCTTG 1629
```

1

```
OM of: US-09-126-945B-2_COPY_21_29.rng * [ ] *

Date: Sep 14, 2001  9:47 PM

About: Results were produced by [illegible]
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL frame+_p2n.model -DEV xlp
 -Q="/cgn2_16/prod/us09_prod/9126945/us-09-126-945b-2_copy_21_29.rng" -DB=N_Geneseq_0601 -QFMT=tastap -SUFFIX=olip2n.rin -GAPOP 4.500
 -GAPEXT 0.050 -MINMATCH=0.100 -100PCL 0.000 -100PEXT 0.000
 QGAPOP 4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT 60.000
 PGAPOP 6.000 -PGAPEXT=7.000 -YGAPOP 60.000 -YGAPEXT 60.000
 -DELOP=6.000 -DELEXT=7.000 -START 1 -MATRIX olig2
 -TRANS=human40.cdi -LIST 45 -DOCALIGN=200 -THR_SCORE quality
 -THR_MIN=1 -ALIGN 15 -MODE=LOCAL -OUTFMT pts -NORM ext -MINLEN 0
 -MAXLEN 2000000000 -USER US09126945_@CGN1_1_408 -NCPU 6 -ICPU 0
 -LONGLOG -NO_XLPXY -WAIT -THREADS 1

Search information block:
Query: US-09-126-945B-2_COPY_21_29
Query length: 9
Database: N_Geneseq_0601:*
Database sequences: 740101
Database length: 313950809
Search time (sec): 680.150000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000


score_list:
Sequence                Strd Orig    %Score   %Score/len  % Recognition ...
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAZ50691 +     9.00   157.68   0.8212     1894  1
/SIDSI/seqdata/geneseq/geneseqn/NA2001.DAT:AAC83261 +     9.00   157.68   0.8212     1894  1
/SIDSI/seqdata/geneseq/geneseqn/NA2001.DAT:AAC83260 +     9.00   153.86   1.44       3817  1
 SIDSI/seqdata/geneseq/geneseqn/NA2002.DAT:AAZ37865 +     7.00   119.40   111.29     2445  1
/SIDSI/seqdata/geneseq/geneseqn/NA1999.DAT:AAY73905 +     7.00   101.79   2.1e-003   32207
/SIDSI/seqdata/geneseq/geneseqn/NA1998.DAT:AAV19041 +     7.00    91.89   3.8e-003   137507
/SIDSI/seqdata/geneseq/geneseqn/NA1996.DAT:AAP30045 +     6.00   110.07   317.67     128  1
/SIDSI/seqdata/geneseq/geneseqn/NA1999.DAT:AAC11489       6.00   107.56   [illegible]  [illegible]  1
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAC05027 +     6.00   116.00   [illegible]  241  1
/SIDSI/seqdata/geneseq/geneseqn/NA1998.DAT:AAV19071 +     6.00   110.51   [illegible]  [illegible]  1
/SIDSI/seqdata/geneseq/geneseqn/NA1998.DAT:AAV19027 +     6.00   110.48   [illegible]  [illegible]  1
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAC01786 +     6.00   110.38   [illegible]  [illegible]  1
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAC01507 +     6.00   110.13   [illegible]  [illegible]  1
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAA15411 +     6.00   110.06   [illegible]  [illegible]  1
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAA08159       6.00   110.47   [illegible]  [illegible]  1
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAA17221 +     6.00   110.62   [illegible]  [illegible]  1
 SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAC08425 +     6.00   111.66   [illegible]  [illegible]  1
/SIDSI/seqdata/geneseq/geneseqn/NA1999.DAT:AAY67610 +     6.00   111.94   [illegible]  [illegible]  1
/SIDSI/seqdata/geneseq/geneseqn/NA1999.DAT:AAY61637 +     6.00   111.94   [illegible]  [illegible]  1
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAC93605 +     6.00   111.28   [illegible]  [illegible]  1
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAC65140 +     6.00   110.25   [illegible]  [illegible]  1
/SIDSI/seqdata/geneseq/geneseqn/NA1999.DAT:AAP37845 +     6.00   110.06   [illegible]  [illegible]  1
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAP38556       6.00   110.01   [illegible]  [illegible]  1
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAP12804 +     6.00   110.08   [illegible]  636  1
/SIDSI/seqdata/geneseq/geneseqn/NA1998.DAT:AAV21307 +     6.00   110.03   [illegible]  [illegible]  1
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAC14855 +     6.00   109.98   [illegible]  [illegible]  1
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAC12670 +     6.00   109.88   [illegible]  [illegible]  1
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAA09130 +     6.00   109.75   [illegible]  [illegible]  1
/SIDSI/seqdata/geneseq/geneseqn/NA1996.DAT:AAY00944 +     6.00   109.50   [illegible]  [illegible]  1
/SIDSI/seqdata/geneseq/geneseqn/NA1996.DAT:AAY65205 +     6.00   109.50   [illegible]  [illegible]  1
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAC16158 +     6.00   109.26   [illegible]  [illegible]  1
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAC22353 +     6.00   108.25   [illegible]  [illegible]  1
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAC46844 +     6.00   108.13   [illegible]  [illegible]  1
/SIDSI/seqdata/geneseq/geneseqn/NA1999.DAT:AAY41976 +     6.00   107.94   [illegible]  [illegible]  1
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAC11550 +     6.00   107.74   [illegible]  [illegible]  1
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAA59889 +     6.00   107.74   [illegible]  [illegible]  1
/SIDSI/seqdata/geneseq/geneseqn/NA2001.DAT:AAE72065 +     6.00   107.47   [illegible]  [illegible]  1
/SIDSI/seqdata/geneseq/geneseqn/NA1996.DAT:AAY65223 +     6.00   106.16   [illegible]  [illegible]  1
/SIDSI/seqdata/geneseq/geneseqn/NA1996.DAT:AAY65224 +     6.00   106.16   [illegible]  [illegible]  1
/SIDSI/seqdata/geneseq/geneseqn/NA1996.DAT:AAY65222 +     6.00   106.16   [illegible]  [illegible]  1
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAZ50692 +     6.00   105.98   [illegible]  [illegible]  1
/SIDSI/seqdata/geneseq/geneseqn/NA1999.DAT:AAV63423 +     6.00   105.98   [illegible]  [illegible]  1
/SIDSI/seqdata/geneseq/geneseqn/NA1999.DAT:AAQ05162 +     6.00   105.59   [illegible]  [illegible]  1
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAC74295       6.00   105.51   661.15     1280  1
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAC41974 +     6.00   105.43   662.51     1285  1
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAC59943       6.00   105.25   680.09     1422  1

seq_name: /SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAZ50691

seq_format: fasta_t  R:
ID   AAZ50691 standard; cDNA; 1894 BP.
XX
AC   AAZ50691;
XX
DT   23-MAY-2000  (first entry)
XX
DE   Human prostate derived Ets factor cDNA
XX
KW   Prostate derived Ets factor; PDEF; human; chromosome 6p21.3; cancer;
KW   loss of heterozygosity; chromosomal translocation; linkage analysis;
KW   cytostatic, cardiant, immunosuppressive, cerebroprotective; fungicide;
KW   antibacterial; vulnerary; neuroprotective; antiparkinsonian; nootropic;
KW   anabolic; antiinflammatory; anorectic; hybridisation probe; forensic;
KW   tumour marker; diagnosis; treatment; prostate cancer; blood coagulation;
KW   autoimmune disorder; haematopoietic; immune/nervous system; stroke;
KW   neoplasm; microbial infection; tissue regeneration; heart attack;
KW   scarring; food additive; preservative; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             416..1423
FT                   /*tag= a
FT                   /product= "Human prostate derived Ets factor"
FT                   /note= "Expressed in prostate epithelium"
XX
PN   WO200006589-A1.
XX
PD   10-FEB-2000.
XX
PF   02-AUG-1999;  99WO-US17470.
XX
PR   31-JUL-1998;  98US-0126945.
XX
PA   (HUMA ) HUMAN GENOME SCI INC.
PA   (PFIB-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI   Liberman JA,  Petty JP,  Eunson CA,  Endress GA,  Rosen CA;
XX
DR   WPI; 2000-195755/17.
DR   P-PSDB; AAY45063.
XX
PT   Novel prostate derived polypeptide, polynucleotide useful for
PT   diagnosis, prevention and treatment of prostate cancer, autoimmune
PT   disorders, microbial infections and also as food additive or
PT   preservative -
XX
PS   Claim 4; Fig 1; 142pp; English.
XX
CC   The present DNA sequence encodes a human prostate derived Ets factor
CC   (PDEF), isolated from human prostate carcinoma cDNA library
CC   (A****** & ******2). PDEF gene expression is highest in tissues involved in
CC   acidic pH sensitivity, with enriched expression in prostate epithelium
CC   cells. PDEF gene is mapped to the human chromosome 6p21.3
CC   region that is associated with loss of heterozygosity and chromosomal
CC   translocations in various human cancers. PDEF has cytostatic, cardiant,
CC   immunosuppressive, cerebroprotective, fungicide, antibacterial,
CC   vulnerary, neuroprotective, antiparkinsonian, nootropic, anabolic,
CC   antiinflammatory and anorectic activity. PDEF polynucleotides are useful
CC   in linkage analysis as markers, as hybridisation probes for differential
CC   identification of the tissues or cell types and as polymorphic markers
CC   for forensic purposes. PDEF is useful as prostate specific tumour marker
CC   for the diagnosis and treatment of prostate cancer. PDEF sequences are
```

CC useful for treating autoimmune disorders, haematopoietic, blood
CC coagulation, immune and nervous system disorders, hyperproliferative
CC disorders like, neoplasms and microbial infections, heart attacks,
CC stroke, scarring and for tissue regeneration. They are also useful as
CC food additives or preservatives.
XX
SQ Sequence 1894 BP; 368 A; 654 C; 571 G; 302 T; 0 other;

alignment scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2_COPY_21_29 x AAZ50691 ..

Align seq 1/1 to: AAZ50691 from: 1 to: 1894

1 AspThrValSerArgThrGlyLeuGln 9
||||||||||||||||||||||||||||
476 GACACGGTGTCGCGGACAGGCTTGGAG 502

Seq name: /cgi6/ptodata/2/pubseq/seqcore/NA2000.ofr:AAZ50691

Seq documentation block:
ID AAZ50691 standard; DNA; 1894 BP.
XX
AC AAZ50691;
XX
DT 16-MAR-2001 (first entry)
XX
DE Gland specific Ets transcription factor (GSEF) nucleotide sequence.
XX
KW Transcription factor; gland specific Ets transcription factor; GSEF;
KW metastatic potential; cancer; tumour; metastasis; breast; prostate;
KW leukaemia; lymphoma; sarcoma; melanoma; chromosome 6p21.1 6p21.3; ds.
XX
OS Homo sapiens.
XX
PN WO200070092-A1.
XX
PD 23-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US13173.
XX
PR 14-MAY-1999; 99US-0134112.
XX
PA (CHIR) CHIRON CORP.
XX
PI Kaufmann J, Xin H, Harrowe G;
XX
DR WPI; 2001-041019/05.
DR P-PSDB; AAB49628.
XX
PT Detecting metastatic and potential metastatic cancerous cells, useful
PT for diagnosing, prognosing, grading and staging of cancers by detecting
PT gland specific Ets transcription factor gene product in a biological
PT sample from a cell
XX
PS Disclosure; Page 81-84; 95pp; English.
XX
CC This invention relates to a method for the detection and determination of
CC the metastatic potential of a cell. The method comprises detecting a
CC gland specific Ets transcription factor (GSEF) gene product in a test
CC sample. Detection of a GSEF gene product in the test sample in amount
CC lower than that in a normal cell, is indicative of a cell with high
CC metastatic potential. The method is useful for determining the metastatic
CC potential of a cell, for the diagnosis and prognosis of cancer as well as
CC grading and staging of cancers by detecting GSEF expression in a
CC biological test sample. The method may also be used to monitor patients
CC having a predisposition to develop a particular cancer. GSEF polypeptides
CC are useful for producing antibodies, in cancer diagnosis, prognosis,
CC grading, staging and management of breast and prostate tumours, and in
CC detecting polymorphisms in the sequence. GSEF genes and proteins are also
CC useful in gene therapy. GSEF gene product expression levels can be used
CC in conjunction with any tissue in which an alteration in GSEF gene
CC product expression levels is associated with development of a
CC cancer associated phenotype. Cancers, which can be monitored include
CC cancers of the prostate, cervix, lung and colon, melanomas, colorectal
CC adenocarcinomas, Wilm's tumour, retinoblastoma, sarcomas, myosarcomas,
CC lung carcinomas, leukaemia, and lymphomas. The GSEF gene is located on
CC human chromosome 6, specifically at 6p21.1-6p21.3. The present sequence
CC represents the DNA encoding GSEF.
XX
SQ Sequence 1894 BP; 368 A; 654 C; 571 G; 302 T; 0 other;

alignment scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2_COPY_21_29 x AAC83261 ..

Align seq 1/1 to: AAC83261 from: 1 to: 1894

1 AspThrValSerArgThrGlyLeuGln 9
||||||||||||||||||||||||||||
476 GACACGGTGTCGCGGACAGGCTTGGAG 502

Seq name: /cgi6/ptodata/2/pubseq/seqcore/NA2000.ofr:AAC83261

Seq documentation block:
ID AAC83261 standard; DNA; 1894 BP.
XX
AC AAC83261;
XX
DT 16-MAR-2001 (first entry)
XX
DE Gland specific Ets transcription factor (GSEF) cDNA sequence.
XX
KW Transcription factor; gland specific Ets transcription factor; GSEF;
KW metastatic potential; cancer; tumour; metastasis; breast; prostate;
KW leukaemia; lymphoma; sarcoma; melanoma; chromosome 6p21.1 6p21.3; ss.
XX
OS Homo sapiens.
XX
PN WO200070092-A1.
XX
PD 23-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US13173.
XX
PR 14-MAY-1999; 99US-0134112.
XX
PA (CHIR) CHIRON CORP.
XX
PI Kaufmann J, Xin H, Harrowe G;
XX
DR WPI; 2001-041019/05.
DR P-PSDB; AAB49633.
XX
PT Detecting metastatic and potential metastatic cancerous cells, useful
PT for diagnosing, prognosing, grading and staging of cancers by detecting
PT gland specific Ets transcription factor gene product in a biological
PT sample from a cell
XX
PS Disclosure; Fig 2; 95pp; English.
XX
CC This invention relates to a method for the detection and determination of
CC the metastatic potential of a cell. The method comprises detecting a
CC gland specific Ets transcription factor (GSEF) gene product in a test
CC sample. Detection of a GSEF gene product in the test sample in amount
CC lower than that in a normal cell, is indicative of a cell with high

CC metastatic potential. The method is useful for determining the metastatic
CC potential of a cell, for the diagnosis and prognosis of cancer as well as
CC grading and staging of cancers by detecting GSEF expression in a
CC biological test sample. The method may also be used to monitor patients
CC having a predisposition to develop a particular cancer. GSEF polypeptides
CC are useful for producing antibodies, in cancer diagnosis, prognosis,
CC grading, staging and management of breast and prostate tumours, and in
CC detecting polymorphisms in the sequence. GSEF genes and proteins are also
CC useful in gene therapy. GSEF gene product expression levels can be used
CC in conjunction with any tissue in which an alteration in GSEF gene
CC product expression levels is associated with development of a
CC cancer-associated phenotype. Cancers, which can be monitored include
CC cancers of the prostate, cervix, lung and colon, melanomas, colorectal
CC adenocarcinomas, Wilm's tumour, retinoblastoma, sarcomas, myosarcomas,
CC lung carcinomas, leukaemia, and lymphomas. The GSEF gene is located on
CC human chromosome 6, specifically at 6p21.1-6p21.3. The present sequence
CC represents the cDNA encoding GSEF.
XX
SQ Sequence 3417 BP; 710 A; 1026 C; 970 G; 711 T; 0 other;

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_21_29 x AAC83266

Align Seq 1/1 to: AAC83266 from: 1 to: 3417

1 AspThrValSerArgThrGlyLeuGlu 9
||||||||||||||||||||||||||||
1875 GACACGGTGTCGCGGACAGGCTTGGAG 1901

seq_name: 'SIDSL:seqdir/genseq1.prodseqDNA??' 15? AAC?????

seq_documentation_block:
ID AAZ37865 standard; DNA; 2435 BP.
XX
AC AAZ37865;
XX
DT 15-FEB-2000 (first entry)
XX
DE Beta glucosidase (Bgl A) gene.
XX
KW Bgl A; beta glucosidase; cellulase; cellulosic biomass; ss.
XX
OS Orpinomyces sp.
XX
FH Key Location/Qualifiers
FT CDS 39..2012
FT /*tag= a
FT /product= BglA
FT /note= "Beta glucosidase"
FT sig_peptide 39..86
FT /*tag= b
FT /note= "putative signal sequence"
FT mat_peptide 87..2009
FT /*tag= c
FT /product= beta_glucosidase
XX
PN WO9957287-A2.
XX
PD 11-NOV-1999.
XX
PF 06-MAY-1999; 99WO-US10106.
XX
PR 06-MAY-1998; 98US-0084494.
XX
PA (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI Li X, Ljungdahl LG, Chen H, Ximenes EA;

XX
DR WPI; 2000-024580/02.
DR P-PSDB; AAY52193.
XX
PT New orpinomyces sp. 2 beta glucosidase coding sequences and protein
XX
PS Claim 2; Page 35-36; 63pp; English.
XX
CC This is the beta glucosidase gene BglA from orpinomyces sp. strain PC-2.
CC Beta-glucosidase acts to liberate D-glucose units from cellobiose,
CC cello-oligosaccharides and other glucosides. A host cell containing the
CC BglA DNA sequence can be used to produce a beta glucosidase in a host
CC cell other than orpinomyces sp strain PC-2. The fungal beta glucosidase
CC is useful for the improved saccharification of cellulosic materials.
CC Effective utilization of cellulosic biomass through biological processes
CC is one approach to overcoming the shortage of food, feeds and fuels,
CC expected as a consequence of the explosive increase in human population.
XX
SQ Sequence 2435 BP; 822 A; 424 C; 452 G; 737 T; 0 other;

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_21_29 x AAZ37865 ..

Align Seq 1/1 to: AAZ37865 from: 1 to: 2435

2 ThrValSerArgThrGlyLeu 8
|||||||||||||||||||||||
257 ACGTCGCGGACAGGCTTGGAGTTA 280

seq_name: 'SIDSL:seqdir/genseq1.prodseqDNA1999.DAT:AAV73805'

seq_documentation_block:
ID AAV73805 standard; DNA; 32207 BP.
XX
AC AAV73805;
XX
DT 25-FEB-1999 (first entry)
XX
DE KSHV LUR DNA (nucleotides 105,401-137,507).
XX
KW Kaposi's sarcoma; acquired immune deficiency syndrome; AIDS; PBPK; Bcl 2;
KW dihydrofolate reductase; LUR; long unique region; vaccine; prophylaxis;
KW diagnosis; treatment; HHV8; capsid protein IV; tegument protein IV;
KW glycoprotein; Kaposin; cyclin D; immediate early protein; IEP; ORF 2;
KW G-adh; G protein coupled receptor; PCAPAI; ds.
XX
OS Kaposi's sarcoma-associated herpesvirus
XX
PN US5849564-A.
XX
PD 15-DEC-1998.
XX
PF 29-NOV-1996; 96US-0770379.
XX
PR 29-NOV-1996; 96US-0770379.
XX
PA (UYSC-) UNIV COLUMBIA NEW YORK.
XX
PI Bohenzky RA, Chang Y, Edelman IS, Moore PS, Rosse LI;
XX
DR WPI; 1999-069741/06.
XX
PI Kaposi's sarcoma-associated herpes virus nucleic acid encodes
PI dihydrofolate reductase and is useful for treatment, prophylaxis
PI or diagnosis of Kaposi's sarcoma
XX
PS Disclosure; Column 155-182; 109pp; English.

```
XX
CC   This sequence is a fragment of the Kaposi's sarcoma associated
CC   herpesvirus (KSHV) LUR (long unique region). This fragment contains
CC   coding regions for ORF65 which encodes capsid protein IV, ORF66, ORF67
CC   which encodes tegument protein IV, ORF68 which encodes a glycoprotein,
CC   ORF69, K12 which encodes kaposin, K13, ORF72 which encodes cyclin D,
CC   ORF73 which encodes immediate early protein (IEP), K14 which encodes
CC   OX 2 (v adh), ORF74 which encodes G protein coupled receptor, ORF75
CC   which encodes tegument protein/FGARAT, K15. KSHV is a new human
CC   Herpesvirus (HHV8) believed to cause Kaposi's sarcoma (KS) which is the
CC   most common form of neoplasm occurring in persons with acquired immune
CC   deficiency syndrome (AIDS). The DHFR protein is useful for vaccination,
CC   prophylaxis, diagnosis and treatment of a subject with Kaposi's sarcoma
CC   and for detecting expression of a DNA virus associated with Kaposi's
CC   Sarcoma in a cell.
XX
SQ   Sequence 32207 BP; 7229 A; 9156 C; 8713 G; 7109 T; 0 other;


alignment_scores.
            Quality:     7.00              Length:       7
              Ratio:  1.000                Gaps:         0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_21_29 x AAV73805/rev

 Align seq 1/1  to reverse of: AAV73805  from: 1  to: 32207

       2 ThrValSerArgThrGlyLeu 8
         ||||||||||||||||||||
    1649 ACTGTCTCACGGACGGGCTTG 1629

seq name: /SIDS1/projects/seqdb/seqdb_NA/98.DA/AAV19941

seq documentation_block:
ID   AAV19941 standard; DNA; 137507 BP.
XX
AC   AAV19941;
XX
DT   04-AUG-1998  (first entry)
XX
DE   KSHV long unique coding region and terminal repeat.
XX
KW   KSHV; HHV8; human herpes virus 8; macrophage inflammatory protein II;
KW   interleukin-6; IL-6; interferon regulatory factor; rheumatoid arthritis;
KW   complement-binding protein; glycoprotein; capsid protein IV; infection;
KW   immediate early protein; Kaposi's sarcoma; protective vaccine; lymphoma;
KW   lymphoproliferative disease; leukaemia; splenomegaly; mycosis fungoides;
KW   HIV immune status; anti inflammatory agent; therapy; ds.
XX
OS   Kaposi's sarcoma associated herpes virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1142..2794
FT                   /*tag   a
FT                   /product  complement-binding protein
FT   CDS             8699..11246
FT                   /*tag   b
FT                   /product  glycoprotein B
FT   CDS             complement (17261..17875)
FT                   /*tag   c
FT                   /product  interleukin 6
FT   CDS             complement (21548..21832)
FT                   /*tag   d
FT                   /product  macrophage inflammatory protein II
FT   CDS             complement (27137..27424)
FT                   /*tag   e
FT                   /product  interferon regulatory factor 1
FT   CDS             28661..29741
FT                   /*tag   f
FT                   /product  protein FIL1
FT   CDS             complement (58976..60175)
FT                   /*tag   g
FT                   /product  glycoprotein M
FT   CDS             complement (69412..69915)
FT                   /*tag   h
FT                   /product  glycoprotein L
FT   CDS             complement (88410..88910)
FT                   /*tag   i
FT                   /product  interferon regulatory factor 2
FT   CDS             89600..90541
FT                   /*tag   j
FT                   /product  interferon regulatory factor 3
FT   CDS             90174..90643
FT                   /*tag   k
FT                   /product  glycoprotein K
FT   CDS             complement (93646..94127)
FT                   /*tag   l
FT                   /product  interferon regulatory factor 4
FT   CDS             complement (111931..112443)
FT                   /*tag   m
FT                   /product  capsid protein IV
FT   CDS             complement (123808..127296)
FT                   /*tag   n
FT                   /product  immediate early protein
XX
PN   WO9804576-A1.
XX
PD   05-FEB-1998.
XX
PF   22-JUL-1997;    97WO-US13346.
XX
PR   29-NOV-1996;    96US-0757669.
PR   25-JUL-1996;    96US-0686244.
PR   25-JUL-1996;    96US-0686349.
PR   25-JUL-1996;    96US-0686350.
PR   25-JUL-1996;    96US-0687254.
PR   25-JUL-1996;    96US-0688814.
PR   05-SEP-1996;    96US-0708678.
PR   10-OCT-1996;    96US-0728324.
PR   13-NOV-1996;    96US-0747887.
PR   13-NOV-1996;    96US-0748640.
XX
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI   Bohenzky RA,  Chang Y,  Edelman IS,  Moore PS,  Russo JJ;
XX
DR   WPI; 1998-130615/12.
XX
PT   New nucleic acid encoding Kaposi's sarcoma associated herpes virus
PT   proteins  useful for, e.g. detecting levels of HHV8 in, and
PT   preparation of vaccines for treatment of, HIV patients
XX
PS   Example 2; Page 145-203; 240pp; English.
XX
CC   This sequence represents the long unique region and terminal repeat of
CC   the Kaposi's sarcoma associated herpes virus (KSHV).  KSHV is also known
CC   as human herpes virus 8 (HHV8). This sequence contains the DNAs of the
CC   invention which encode KSHV polypeptides selected from: (a) viral
CC   macrophage inflammatory protein (MIP) II; (b) viral interleukin 6 (IL-6);
CC   (c) viral IRF 1; (d) complement-binding protein; glycoproteins B, M or L;
CC   (d) capsid protein IV encoded by ORF65; and (e) immediate early protein
CC   encoded by ORF73. Labelled probes for the nucleic acid, proteins encoded
CC   by it, and antibodies (Ab) specific for the proteins are useful for
CC   detecting HHV8, specifically for diagnosis of Kaposi's sarcoma, in body
CC   fluids or tissue samples. HHV8 infections can be treated with antisense
CC   or triplex forming molecules or agents that bind specifically to the
CC   protein. Ab may be used for prophylaxis or treatment of HHV8 infection,
CC   while the protein can be used in protective vaccines. Ab may also be used
CC   to differentiate between lymphomas, and HHV8 may be implicated in many
CC   other lymphoproliferative diseases such as lymphomas, leukaemia,
CC   splenomegaly and mycosis fungoides. Cells and animals containing the
CC   nucleic acid are useful for drug screening. HHV8 derived peptides can be
CC   used as targets for antiviral drugs, e.g. dihydrofolate reductase gene
CC   can be inhibited with methotrexate. These can also be used to determine
```

```
XX   the immune status of a patient infected with HIV. HHV8 derived protein
CC   viral MIP III may be used as an anti-inflammatory agent for,
CC   e.g. treating rheumatoid arthritis. This sequence is stated as containing
CC   81 open reading frames.
XX
SQ   Sequence 147587 BP; 32579 A; 47795 C; 45758 G; 21475 T; 0 other.


alignment_scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_21_29 x AAV19941/fra    .

 Align Seq 1/2   to:  [illegible]  AAV19941  from:  1  to:  147587

       2 ThrValSerArgThrGlyLeu 8
         |||||||||||||||||||||
  106949 ACTGTCTCACGGACGGGCTTG 106929

seq_name:  [illegible]

seq_documentation_block:
ID   AAI30945 standard; cDNA; 178 BP.
XX
AC   AAI30945;
XX
DT   06-AUG-1996  (first entry)
XX
DE   Human neuronatin genomic fragment.
XX
KW   Neuronatin-alpha; rat; neuronatin-beta; neuronatin; brain tissue;
KW   neuronal proliferation; neuronal migration; genetic screening; human;
KW   neural regeneration; ss.
XX
OS   Homo sapiens.
XX
PN   WO9534640-A1.
XX
PD   21-DEC-1995.
XX
PF   07-JUN-1995;   95WO-US07403.
XX
PR   10-JUN-1994;   94US-0259299
XX
PA   (FORD-) FORD HEALTH SYSTEM HENRY.
XX
PI   Joseph R;
XX
DR   WPI; 1996-049665/05.
XX
PT   New isolated neuronal-neonatal neuronatin gene - used to develop
PT   prods. for screening for genetic defects and for use in neural
PT   regeneration
XX
PS   Claim 14; Page 41; 62pp; English.
XX
CC   This sequence represents a fragment of the genomic human neuronatin
CC   sequence.  Neuronatin is expressed in brain tissue during rapid growth,
CC   when there is a surge in neuronal proliferation and migration.  This
CC   sequence was obtained after screening a human foetal cDNA library with
CC   the rat neuronatin cDNA sequences (see AAI30932 and AAI30933).  The rat
CC   sequences were obtained by extracting total RNA from neonatal, young
CC   adult, and adult rat brain.  The RNA was then processed for differential
CC   display.  A cDNA fragment selectively expressed in neonatal brain was
CC   identified.  Neuronatin can be used for screening foetal tissue for
CC   neuronatin gene defects and for screening for potential carriers of the
CC   defect.  The neuronatin sequences can also be used for neural
CC   regeneration.
XX
SQ   Sequence 178 BP; 31 A; 56 C; 65 G; 26 T; 0 other.
```

```
alignment_scores:
            Quality:     6.00              Length:       6
              Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_21_29 x AAI30935    ..

 Align Seq 1/2   to:  AAI30935  from:  1  to:  178

       2 ThrValSerArgThrGly 7
         | |||||||||||||||
  36 ACGGTGTCGCGGACGCGG 55

Query  [illegible]  Database:  [illegible]  AAZ14488

seq_documentation_block:
ID   AAZ14488 standard; cDNA; 218 BP.
XX
AC   AAZ14488;
XX
DT   12-OCT-1999  (first entry)
XX
DE   Human gene expression product cDNA sequence SEQ ID NO:1957.
XX
KW   Human; gene; gene expression product; diagnosis; therapy; probe;
KW   detection; mapping; tissue typing; profiling; forensic; cancer;
KW   genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO9938972-A2.
XX
PD   05-AUG-1999.
XX
PF   28-JAN-1999;   99WO-US01619.
XX
PR   03-APR-1998;   98US-0080666.
PR   28-JAN-1998;   98US-0072910.
PR   24-FEB-1998;   98US-0075954.
PR   31-MAR-1998;   98US-0080114.
PR   03-APR-1998;   98US-0080515.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Ciruvenjakov B;  Dickson M;  Drmanac R;  Drmanac S;
PI   Escobedo J;  Garcia PD;  Garcia V;  Gieser K;  Innis MA;
PI   Jones WL;  Kassam A;  Kennedy GC;  Kita D;  Labat I;
PI   Lamson G;  Leshkowitz D;  Pal D;  Randazzo F;  Reinhard C;
PI   Stuelpnagel R;  Sudduth Klinger J;  Williams LT;
XX
DR   WPI; 1999-494092/41.
XX
PT   Novel human genes and their expression products which are
PT   differentially expressed in different cell types
XX
PS   Claim 1; Page 1102; 2479pp; English.
XX
CC   The present invention describes a library of human polynucleotides
CC   comprising the sequences given in AAZ12542 to AAZ13779. Also described is
CC   a method of detecting differentially expressed genes correlated with the
CC   cancerous state of a mammalian cell, comprising detecting at least one
CC   differentially expressed gene product in a test sample from a cell
CC   suspected of being cancerous, where the gene product is encoded by one
CC   of the 5248 polynucleotide sequences given in AAZ12542 to AAZ13779. The
CC   polynucleotides can be used as a source of primers and probes, which can
CC   be used for a variety of purpose, e.g. detection of expression levels,
CC   mapping, tissue typing or profiling, forensics, genetic analysis and
CC   detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC   can be used for raising antibodies for experimental, diagnostic and
```

```
CC   therapeutic purposes. The polynucleotides may also be used to construct
CC   arrays for diagnostics (which may be used to determine function of an
CC   encoded protein); and to detect differences in expression levels between
CC   two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC   identify a genetic predisposition or susceptibility to a disease such as
CC   cancer). The polynucleotides of the invention are especially used in the
CC   diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC   and lung cancer. The polynucleotides can also be used to screen for
CC   peptide analogues and antagonists.
XX
SQ   Sequence 218 BP; 46 A; 55 C; 63 G; 47 T; 7 other;


alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_21_29 x AAZ14488/rev    .

 Align seq 1/1  to reverse of: AAZ14488  from: 1  to: 218

       4 SerArgThrGlyLeuGln 9
         |||||||||||||||||
     104 TCCAGGACAGGCTTAGAG 87

Seq name: /cgi-1/seqdata/processed/geneseq/GeneSeqData_*_AA_*.fsa:AAZ14488

Seq documentation block:
ID   AAC05027 standard; cDNA; 241 BP.
XX
AC   AAC05027;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 9102.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 9102; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 99 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 241 BP; 50 A; 94 C; 44 G; 52 T; 1 other;


alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_21_29 x AAC05027    .

 Align seq 1/1  to: AAC05027  from: 1  to: 241

       4 SerArgThrGlyLeuGln 9
         |||||||||||||||||
     181 AGCAGGACAGGCTTGGAA 198

Seq name: /cgi-1/seqdata/processed/geneseq/GeneSeqData_*_DNA:AAV19004

Seq documentation block:
ID   AAV19004 standard; cDNA; 254 BP.
XX
AC   AAV19004;
XX
DT   10-AUG-1998  (first entry)
XX
DE   Human neuronatin nucleotide sequence.
XX
KW   Human neuronatin; HNEU; senescence; neuronal cell; gene therapy;
KW   abnormal brain development; aging; Alzheimer's disease; antibody;
KW   nervous system; Parkinson's disease; immunoassay; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             62..227
FT                   /*tag= a
FT                   /product= "human neuronatin"
XX
PN   WO9804692-A2.
XX
PD   05-FEB-1998.
XX
PF   21-JUL-1997;   97WO-US12735.
XX
PR   31-JUL-1996;   96US-0690094.
XX
PA   (INCY ) INCYTE PHARM INC.
XX
PI   Coleman R, Stuart SG;
XX
DR   WPI; 1998-130684/12.
DR   P-PSDB; AAW37777.
XX
PT   New human neuronatin - used for treatment of abnormal brain
PT   development or senescence, e.g. Alzheimer's or Parkinson's disease,
PT   for diagnosis and for identifying specific modulators
XX
PS   Claim 3; Page 38; 46pp; English.
XX
CC   This nucleotide sequence encodes the novel human neuronatin (HNEU).
CC   Reduction in HNEU expression is associated with senescence in neuronal
CC   cells. It may be optionally expressed from gene therapy vectors, to
CC   treat abnormal brain development or aging, e.g. Alzheimer's disease,
CC   also disorders of the central or peripheral nervous systems, e.g.
CC   Parkinson's disease. HNEU antibodies are used as immunoassay reagents
CC   for diagnosis and monitoring HNEU associated disease.
XX
```

```
SQ   Sequence 254 BP; 42 A; 87 C; 80 G; 45 T; 0 other;


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_21_29 x AAV19004    ..

 Align seq 1/1  to: AAV19004  from: 1  to: 254

        2 ThrValSerArgThrGly 7
          ||||||||||||||||||
      168 ACGGTGTCGCGGACCGGG 185

Seq_name: /SIDS1/bcrilata/ps_seq/pctoseq_pc/NAC_c_DAT_AAV19006.

Seq_documentation_block:
ID   AAV19006 standard; cDNA; 257 BP.
XX
AC   AAV19006;
XX
DT   10-AUG-1998  (first entry)
XX
DE   Human neuronatin incyte clone 926572.
XX
KW   Human neuronatin; HNEU; clone 926572; senescence; neuronal cell;
KW   gene therapy; abnormal brain development; ataxia; Alzheimer's disease;
KW   antibody; nervous system; Parkinson's disease; immunoassay; ss.
XX
OS   Homo sapiens.
XX
PN   WO9804692 A2.
XX
PD   05-FEB-1998.
XX
PF   21-JUL-1997;   97WO-US12735.
XX
PR   31-JUL-1996;   96US-0690094.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Coleman R, Stuart SG;
XX
DR   WPI; 1998-130684/12.
XX
PT   New human neuronatin - used for treatment of abnormal brain
PT   development or senescence, e.g. Alzheimer's or Parkinson's disease,
PT   for diagnosis and for identifying specific modulators
XX
PS   Disclosure; Page 38; 46pp; English.
XX
CC   This is the nucleotide sequence of the human neuronatin (HNEU)
CC   incyte clone 926572.  Reduction in HNEU expression is associated
CC   with senescence in neuronal cells.  It may be optionally expressed
CC   from gene therapy vectors, to treat abnormal brain development or
CC   ataxia, e.g. Alzheimer's disease, also disorders of the central or
CC   peripheral nervous systems, e.g. Parkinson's disease.  HNEU
CC   antibodies are used as immunoassay reagents for diagnosis and
CC   monitoring HNEU-associated disease.
XX
SQ   Sequence 257 BP; 43 A; 68 C; 82 G; 45 T; 19 other;


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_21_29 x AAV19006    ..

 Align seq 1/1  to: AAV19006  from: 1  to: 257

        2 ThrValSerArgThrGly 7
          ||||||||||||||||||
      174 ACGGTGTCGCGGACCGGG 191

Seq_name: /SIDS1/bcrilata/ps_seq/pctoseq_pc/NAC_c_DAT_AAC01786.

Seq_documentation_block:
ID   AAC01786 standard; cDNA; 301 BP.
XX
AC   AAC01786;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 1784.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401 A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99GB-0122487.
XX
PA   (GESI ) GENSET.
XX
PI   Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR   WPI; 2000-566381/45.
DR   P-PSDB; AAC01786.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 1784; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. An ORF has been identified within the
CC   sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC   derived from 30 different tissues. EST sequences usually correspond
CC   mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC   often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC   well suited for isolating cDNA sequences derived from the 5' ends of
CC   mRNAs and even in those cases where longer cDNA sequences have been
CC   obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC   mRNAs with intact 5' ends and can therefore be used to obtain full length
CC   cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC   gene therapy and chromosome mapping procedures. They are used to obtain
CC   upstream regulatory sequences and to design expression and secretion
CC   vectors.
XX
SQ   Sequence 301 BP; 96 A; 63 C; 72 G; 70 T; 0 other;


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_21_29 x AAC01786    ..

 Align seq 1/1  to: AAC01786  from: 1  to: 301

        2 ThrValSerArgThrGly 7
          ||||||||||||||||||
      174 ACGGTGTCGCGGACCGGG 191
```

```
          ||||||||||||||||||
       17 GTCTCCTAGCACTGCACTC 34

Seq name: /cis03/ftp/data/geneseq/geneseqp/NA2000.DAT:AAC25027

Seq documentation block:
ID   AAC25027 standard; cDNA; 414 BP.
XX
AC   AAC25027;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 29102.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;  99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR   WPI; 2000-500481/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures
XX
PS   Claim 1; SEQ ID 29102; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 40 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 414 BP; 62 A; 72 C; 99 G; 75 T; 6 other;


Alignment scores:
             Quality:     6.00              Length:       6
               Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

Alignment block:
 US-09-126-945B-2-COPY_21_29 x AAC25027  ..

 Align seq 1/1  to: AAC25027  from: 1  to: 414

        8 ValSerArgThrGlyLeu 8
          ||||||||||||||||||
      208 GTGTCCAGAACTGCACTG 225

Seq name: /cis03/ftp/data/geneseq/geneseqp/NA2000.DAT:AAA15448

Seq documentation block:
ID   AAA15448 standard; DNA; 330 BP.
XX
AC   AAA15448;
XX
DT   04-SEP-2000  (first entry)
XX
DE   DNA sequence whose expression is regulated by presence of opiates.
XX
KW   opiate; haematopoietic cell; immune deficiency; blood protein disorder;
KW   agammaglobulinemia; dysgammaglobulinemia; ataxia telangiectasia;
KW   common variable immunodeficiency; Digeorge Syndrome; HIV infection;
KW   HTLV; Human T-cell Leukaemia; Lymphoma virus infection;
KW   leucocyte adhesion deficiency syndrome; lymphopenia; haemoglobinuria;
KW   phagocyte bactericidal dysfunction; severe combined immunodeficiency;
KW   SCID; Wiskott-Aldrich disorder; anemia; thrombocytopenia; ss.
XX
OS   Mus musculus.
XX
PN   WO200056876-A2.
XX
PD   28-MAY-2000.
XX
PF   19-OCT-1999;   99WO-US24708.
XX
PR   10-NOV-1998;   98US-0107806.
XX
PA   (DIGI ) DIGITAL GENE TECHNOLOGIES INC.
XX
PI   Sutcliffe JG, Przewlocki R, Quan J, Thomas EA;
XX
DR   WPI; 2000-379541/32.
XX
PT   New polynucleotides whose expression is regulated by opiate
PT   administration and withdrawal, useful for treating blood protein
PT   disorders, ataxia telangiectasia and common variable immunodeficiency
XX
PS   Claim 3; Page 83; 85pp; English.
XX
CC   AAA15419-58 represent polynucleotides whose expression is regulated by
CC   opiate administration and withdrawal. The polynucleotides are useful in
CC   methods for preventing, treating or ameliorating a medical condition.
CC   Diagnostic methods are used to detect mutations in the polynucleotides.
CC   They could be used to increase differentiation and proliferation of
CC   haematopoietic cells, including the pluripotent stem cells, in an effort
CC   to treat those disorders associated with a decrease in certain (or
CC   many) types haematopoietic cells.  Examples of immunologic deficiency
CC   syndromes include, but are not limited to blood protein disorders (e.g.
CC   agammaglobulinemia, dysgammaglobulinemia), ataxia telangiectasia,
CC   common variable immunodeficiency, Digeorge Syndrome, HIV infection,
CC   HTLV (human T-cell leukaemia/ lymphoma virus ) BLV infection, leucocyte
CC   adhesion deficiency syndrome, lymphopenia, phagocyte bactericidal
CC   dysfunction, severe combined immunodeficiency (SCIDs), Wiskott-Aldrich
CC   disorder, anemia, thrombocytopenia, or haemoglobinuria.
XX
SQ   Sequence 330 BP; 96 A; 64 C; 78 G; 90 T; 2 other;


Alignment scores:
             Quality:     6.00              Length:       6
               Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

Alignment block:
 US-09-126-945B-2_COPY_21_29 x AAA15448  ..

 Align seq 1/1  to: AAA15448  from: 1  to: 330

        1 AspThrValSerArgThr 6
          ||||||||||||||||||
      101 GACACTGTCTCAAGAACA 118

Seq name: /cis03/ftp/data/geneseq/geneseqp/NA2000.DAT:AAA68150
```

```
seq_documentation_block:
ID   AAA68150 standard; DNA; 344 BP.
XX
AC   AAA68150;
XX
DT   24-OCT-2000  (first entry)
XX
DE   Eucalyptus grandis PAL nucleotide sequence SEQ ID NO:243.
XX
KW   Plant; lignin; lignin biosynthetic pathway; Eucalyptus grandis;
KW   Pinus radiata; Monterey pine; ds.
XX
OS   Eucalyptus grandis.
XX
PN   WO200022099-A1.
XX
PD   20-APR-2000.
XX
PF   06-OCT-1999;   99WO-NZ00168.
XX
PR   09-OCT-1998;   98US-0169789.
PR   14-JUL-1999;   99US-0143811.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Bloksberg LN,  Havukkala IJ;
XX
DR   WPI; 2000-317962/27.
XX
PT   Novel polynucleotide encoding enzymes involved in lignin biosynthetic
PT   pathway useful for producing transgenic plants especially eucalyptus
PT   and pine species having altered lignin content, composition and
PT   structure -
XX
PS   Claim 1; Page 131; 213pp; English.
XX
CC   The present invention describes isolated polynucleotides and proteins
CC   encoding and representing the enzymes cinnamate 4-hydroxylase (C4H),
CC   coumarate 3-hydroxylase (C3H), phenolase (PNL), O-methyl transferase
CC   (OMT), cinnamyl alcohol dehydrogenase (CAD), cinnamoyl-CoA reductase
CC   (CCR), phenylalanine ammonia-lyase (PAL), 4-coumarate:CoA ligase (4CL),
CC   coniferol glucosyl transferase (CGT), coniferin beta glucosidase (CBG),
CC   laccase, peroxidase, ferulate-5-hydroxylase (F5H), alpha-amylase,
CC   caffeic acid methyl transferase, caffeoyl CoA methyl transferase,
CC   coumerate CoA ligase, cytochrome P450 LXX1A, diphenol oxidase, flavanol
CC   glucosyl transferase, flavenoid hydroxylase, and isoflavone reductase,
CC   which are involved in the lignin biosynthetic pathway. The
CC   polynucleotides can be used for modulating lignin content, lignin
CC   composition and the structure of a plant, especially eucalyptus and pine
CC   species, and for modifying the activity of an enzyme involved in lignin
CC   biosynthetic pathway, and for producing a plant having altered lignin
CC   content, composition and structure. They can be used for designing probes
CC   and primers useful for detecting similar DNA and RNA sequences in any
CC   organism and for PCR amplification. The lignin content can be efficiently
CC   modified using the polynucleotides. AAA67908 to AAA68201 and AAB16341 to
CC   AAB16449 represent polynucleotide and protein sequences used in the
CC   exemplification of the present invention.
XX
SQ   Sequence 344 BP; 78 A; 90 C; 111 G; 65 T; 0 other;


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_21_29 x AAA68150/rev   ..

 Align seq 1/1  to reverse of: AAA68150  from: 1  to: 344
```

```
  4 SerAsnIleGlyLeuGln 9
    |||| ||||||||||||||
141 TCTCGIACIGGACTCGAA 124
```

OM nt - US-09-126-945B-2_COPY_21_29 [illegible] Issued_Patents_NA:* [illegible]

Date: Sep 14 2001 9:53 PM

About: Results were produced by the [illegible] software [illegible] 4.5
Copyright (c) [illegible] Compugen Ltd.

Command line parameters:
-MODEL frame+_p2n.model -DEV xlp
-Q [illegible] /app_query.fasta_1.1267
-DB Issued_Patents_NA -QFMT fastap -SUFFIX olip2n.rni
-GAPOP 4.500 -GAPEXT 0.050 -MINMATCH 0.100 -LOOPCL 0.000
-LOOPEXT 0.000 -QGAPOP 4.500 -QGAPEXT 0.050 -XGAPOP 60.000
-XGAPEXT 60.000 -FGAPOP 6.000 -FGAPEXT 7.000 -YGAPOP 60.000
-YGAPEXT 60.000 -DELOP 6.000 -DELEXT 7.000 -START 1
-MATRIX oligo -TRANS human40.cdi -LIST 45 -DOCALIGN 200
-THR_SCORE quality -THR_MIN 1 -ALIGN 15 -MODE LOCAL -OUTFMT pto
-NORM ext -MINLEN 0 -MAXLEN 2000000000
-USER US09126945_@CGN1_1_93 -NCPU 6 -ICPU 3 -LONGLOG -NO_XLPY
-WAIT -THREADS 1

Search information block:
Query: US-09-126-945B-2_COPY_21_29
Query length: 9
Database: Issued_Patents_NA:*
Database sequences: 324599
Database length: 94655562
Search time (sec): 422.720000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	! Documentation ..
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-306-593-1 +		7.00	117.80	60.40	2445 !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-551-437-3		7.00	117.80	60.40	3084 !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-004-225-3 -		7.00	117.80	60.40	3084 !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-084-246-3 -		7.00	117.80	60.40	3084 !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-[illegible]		7.00	102.00	458.55	32007 !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-[illegible]A-20 -		7.00	102.00	458.55	32007
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-911-894-71 +		6.00	110.01	164.12	644 !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-602-093-5 +		6.00	106.41	177.26	704 !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-602-092-2 +		6.00	106.32	263.33	1113 !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-602-093-1 +		6.00	105.94	280.01	1165 !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-[illegible]-21 +		6.00	105.70	292.24	[illegible] !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-712-[illegible] +		6.00	103.22	387.65	1762 !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-554-612C-13 -		6.00	103.14	395.87	1784 !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-231-[illegible]		6.00	102.45	432.61	1977 !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-977-816-2 -		6.00	102.45	432.61	1977 !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-421-661-5 +		6.00	100.96	522.46	2465 !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-014-969-1		6.00	100.84	531.52	2509 !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-554-61[illegible]		6.00	[illegible]	[illegible]	[illegible] !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-010-998-1 +		6.00	100.04	589.07	2826 !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-551-6[illegible]-51		6.00	99.87	[illegible]	[illegible] !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-554-612C-10 -		6.00	99.85	603.99	2909 !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-551-61[illegible]-11		6.00	99.85	603.99	2909 !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-[illegible]		6.00	97.75	790.69	3573 !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-554-612C-2		6.00	95.15	1.16e+03	5843 !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-306-691B-42 -		6.00	93.92	1.3e+03	7011 !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-345-[illegible]		6.00	[illegible]	[illegible]	[illegible] !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-336-[illegible]A-1		6.00	77.51	[illegible]+04	[illegible]161 !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-[illegible]A-[illegible]		5.00	116.[illegible]	67.[illegible]	[illegible] !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-[illegible]-7		5.00	115.05	85.41	[illegible] !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-639-[illegible]		5.00	115.05	85.41	[illegible] !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-639-[illegible]		5.00	115.05	85.41	[illegible] !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-[illegible]-91		5.00	115.05	85.41	[illegible] !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-011-[illegible]-91 -		5.00	115.05	85.41	[illegible] !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-044-908-81		5.00	115.05	85.41	[illegible] !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-[illegible]A-121 +		5.00	115.05	85.41	[illegible] !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-071-[illegible]A-127		5.00	[illegible]	101.[illegible]	[illegible] !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-[illegible]-8		5.00	112.55	101.[illegible]	[illegible] !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-17[illegible]-8		5.00	112.55	101.[illegible]	[illegible] !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-125-[illegible]		5.00	112.[illegible]	114.94	2[illegible] !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-074-[illegible]A-7[illegible]		5.00	11[illegible]	122.[illegible]	[illegible] !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-[illegible]-492-9 -		5.00	111.88	129.00	32 !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-906-517-9		5.00	111.88	129.00	32 !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-382-616A-32 +		5.00	111.[illegible]	142.82	[illegible]
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-382-616A-33 -		5.00	111.[illegible]	142.82	[illegible]
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-[illegible]-14		5.00	[illegible]	[illegible].81	41 !

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-306-593-1

seq_documentation_block:
; Sequence 1, Application US/09306593
; Patent No. 6184018
; GENERAL INFORMATION:
; APPLICANT: Li, Xin Liang
; APPLICANT: Ljungdahl, Lars G.
; APPLICANT: Chen, Huizhong
; APPLICANT: Ximenes, Eduardo A.
; TITLE OF INVENTION: Beta-glucosidase Coding Sequences and Protein from
; TITLE OF INVENTION: Orpinomyces PC-2
; FILE REFERENCE: 41-98us
; CURRENT APPLICATION NUMBER: US/09/306,593
; CURRENT FILING DATE: 1999-05-06
; EARLIER APPLICATION NUMBER: US 60/084,494
; EARLIER FILING DATE: 1998-05-06
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2445
; TYPE: DNA
; ORGANISM: Orpinomyces sp. PC-2
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (39)..(2009)
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (87)..(2009)
US-09-306-593-1

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_21_29 x US-09-306-593-1 ..

Align seq 1/1 to: US-09-306-593-1 from: 1 to: 2445

2 HisValSerArgThrGlyLeu 8
|||| ||||||||||||||||
[illegible] A[illegible]CGCAA[illegible]ATTA 777

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-551-437-3

seq_documentation_block:
; Sequence 3, Application US/08551437
; Patent No. 5824545
; GENERAL INFORMATION:
; APPLICANT: Koltin, Yigal
; APPLICANT: Riggle, Perry
; APPLICANT: Gavrias, Vicky
; APPLICANT: Bulawa, Chris
; APPLICANT: Winter, Ben
; TITLE OF INVENTION: [illegible] EUKARYOTIC
; TITLE OF INVENTION: [illegible] GENES AND PROMOTER ISOLATION VECTOR
; TITLE OF INVENTION: AND METHOD OF USE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston

```
;       STATE:  MA
;       COUNTRY:  US
;       ZIP:  02110-2804
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  Windows95
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/551,437
;       FILING DATE:  01-NOV-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Fasse, J. Peter
;       REGISTRATION NUMBER:  32,983
;       REFERENCE/DOCKET NUMBER:  06286/009001
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617/542-507
;       TELEFAX:  617/542-890
;       TELEX:  200154
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3084 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
US-08-551-437-3


alignment_scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block
 US-09-126-945B-2_COPY_21_29 x US-08-551-437-3/rev   ..

 Align seq 1/1  to reverse of: US-08-551-437-3  from: 1  to: 3084

        2 ThrValSerArgThrGlyLeu 8
          ||||||||||||||||||||||
    1258 ACTGTATCCCGGACCGGTTTA 1238

seq_name: /seqn2_6/ptodata/2/ina/6A_COMB.seq:US-09-004-225-3

seq_documentation_block:
; Sequence 3, Application US/09004225
; Patent No. 6020133
;  GENERAL INFORMATION:
;     APPLICANT:  Koltin, Yigal
;     APPLICANT:  Riggle, Perry
;     APPLICANT:  Gavrias, Vicky
;     APPLICANT:  Bulawa, Chris
;     APPLICANT:  Winter, Ben
;     TITLE OF INVENTION:  IDENTIFICATION OF EUKARYOTIC
;     TITLE OF INVENTION:  GROWTH-RELATED GENES AND PROMOTER ISOLATION VECTOR
;     TITLE OF INVENTION:  AND METHOD OF USE
;     NUMBER OF SEQUENCES:  6
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Fish & Richardson, P.C.
;       STREET:  225 Franklin Street
;       CITY:  Boston
;       STATE:  MA
;       COUNTRY:  US
;       ZIP:  02110-2804
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  Windows95
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/004,225
;       FILING DATE:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/551,437
;       FILING DATE:  01-NOV-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Fasse, J. Peter
;       REGISTRATION NUMBER:  32,983
;       REFERENCE/DOCKET NUMBER:  06286/009001
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617/542-507
;       TELEFAX:  617/542-890
;       TELEX:  200154
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3084 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
US-09-004-225-3


alignment_scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block
 US-09-126-945B-2_COPY_21_29 x US-09-004-225-3/rev   ..

 Align seq 1/1  to reverse of: US-09-004-225-3  from: 1  to: 3084

        2 ThrValSerArgThrGlyLeu 8
          ||||||||||||||||||||||
    1258 ACTGTATCCCGGACCGGTTTA 1238

seq_name: /seqn2_6/ptodata/2/ina/6B_COMB.seq:US-09-084-346-3

seq_documentation_block:
; Sequence 3, Application US/09084346
; Patent No. 6251594
;  GENERAL INFORMATION:
;     APPLICANT:  Koltin, Yigal
;     APPLICANT:  Riggle, Perry
;     APPLICANT:  Gavrias, Vicky
;     APPLICANT:  Bulawa, Chris
;     APPLICANT:  Winter, Ben
;     TITLE OF INVENTION:  IDENTIFICATION OF EUKARYOTIC
;     TITLE OF INVENTION:  GROWTH-RELATED GENES AND PROMOTER ISOLATION VECTOR
;     TITLE OF INVENTION:  AND METHOD OF USE
;     NUMBER OF SEQUENCES:  6
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Fish & Richardson, P.C.
;       STREET:  225 Franklin Street
;       CITY:  Boston
;       STATE:  MA
;       COUNTRY:  US
;       ZIP:  02110-2804
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  Windows95
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/084,346
;       FILING DATE:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/551,437
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Fasse, J. Peter
;       REGISTRATION NUMBER:  32,983
;       REFERENCE/DOCKET NUMBER:  06286/009001
```

```
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617/542-5070
;       TELEFAX:  617/542-8906
;       TELEX:  200154
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4084 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
US-09-084-346-3


alignment_scores:
             Quality:    7.00              Length:       7
               Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_21_29 x US-09-084-346-3/rev     .

 Align Seq 1/1  to reverse of: US-09-084-346-3  from: 1  to: 4084

        2 IhrValSerArgIhrGlyLeu 8
          ||||||||||||||||||||||
     1258 ACTCTATCCCGGACCGGTTTA 1238

Seq Name: /tmp/ / pf_data/2/ica/59-COMP_seq/US-09-126-945B-2_COPY_21_29.rni

seq_documentation_block:
; Sequence 20, Application US/08770379
; Patent No. 5849564
;  GENERAL INFORMATION:
;     APPLICANT:  Chang, Yuan
;     APPLICANT:  Bohenzky, Roy A.
;     APPLICANT:  Russo, James J.
;     APPLICANT:  Edelman, Isidore S.
;     APPLICANT:  Moore, Patrick S.
;     TITLE OF INVENTION:  POLYPEPTIDES FROM KAPOSI'S SARCOMA ASSOCIATED
;     TITLE OF INVENTION:  HERPESVIRUS DNA PROTEINS AND USES THEREOF
;     NUMBER OF SEQUENCES:  20
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Cooper & Dunham LLP
;       STREET:  1185 Avenue of the Americas
;       CITY:  New York
;       STATE:  New York
;       COUNTRY:  U.S.A.
;       ZIP:  10036
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/770,379
;       FILING DATE:
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  White, John P.
;       REGISTRATION NUMBER:  28,678
;       REFERENCE/DOCKET NUMBER:  52342
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 278-0400
;       TELEFAX:  (212) 391-0525
;   INFORMATION FOR SEQ ID NO:  20:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  32207 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-770-379-20


alignment_scores:
             Quality:    7.00              Length:       7
               Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_21_29 x US-08-770-379-20/rev     .

 Align Seq 1/1  to reverse of: US-08-770-379-20  from: 1  to: 32207

        2 IhrValSerArgIhrGlyLeu 8
          ||||||||||||||||||||||
     1649 ACTGTCTCAGGACGGGCTTG 1629

Seq Name: /tmp/ / pf_data/2/ica/59-COMP_seq/US-09-126-945B-2_COPY_21_29.rni

seq_documentation_block:
; Sequence 20, Application US/08757669A
; Patent No. 5849751
;  GENERAL INFORMATION:
;     APPLICANT:  Chang, Yuan
;     APPLICANT:  Bohenzky, Roy A.
;     APPLICANT:  Russo, James J.
;     APPLICANT:  Edelman, Isidore S.
;     APPLICANT:  Moore, Patrick S.
;     TITLE OF INVENTION:  UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
;     TITLE OF INVENTION:  SEQUENCES AND USES THEREOF
;     NUMBER OF SEQUENCES:  20
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Cooper & Dunham LLP
;       STREET:  1185 Avenue of the Americas
;       CITY:  New York
;       STATE:  New York
;       COUNTRY:  U.S.A.
;       ZIP:  10036
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/757,669A
;       FILING DATE:
;       CLASSIFICATION:  424
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  White, John P.
;       REGISTRATION NUMBER:  28,678
;       REFERENCE/DOCKET NUMBER:  45185-F
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 278-0400
;       TELEFAX:  (212) 391-0525
;   INFORMATION FOR SEQ ID NO:  20:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  32207 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-757-669A-20


alignment_scores:
             Quality:    7.00              Length:       7
               Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_copy_21_29 x US-08-757-669A-20/rev     .

 Align Seq 1/1  to reverse of: US-08-757-669A-20  from: 1  to: 32207

        2 IhrValSerArgIhrGlyLeu 8
```

```
          ||||||||||||||||||||
    1649 ACTGTCTCACGGACGGGCTTG 1629

seq_name: /cgn2_6/ptodata/2/iia/6A_COMP_seq_US_08_911_894_71

seq_documentation_block:
; Sequence 71, Application US/08911894
; Patent No. 6030840
;  GENERAL INFORMATION:
;     APPLICANT:  Saxon, Andrew
;     APPLICANT:  Zhang, Ke
;     APPLICANT:  Fujieda, Shigeharu
;     TITLE OF INVENTION:  IMMUNOGLOBULIN TRANS-SPLICED TRANSCRIPTS
;     TITLE OF INVENTION:  AND USES THEREOF
;     NUMBER OF SEQUENCES:  90
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Akin, Gump, Strauss, Hauer & Feld
;       STREET:  816 Congress Avenue, Suite 1900
;       CITY:  Austin
;       STATE:  Texas
;       COUNTRY:  USA
;       ZIP:  78701
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/911,894
;       FILING DATE:  Concurrently Herewith
;       CLASSIFICATION:  536
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 60/024,579
;       FILING DATE:  19-AUG-1996
;       CLASSIFICATION:  536
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Mayfield, Denise L.
;       REGISTRATION NUMBER:  33,732
;       REFERENCE/DOCKET NUMBER:  43496.0006
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (512) 499-6200
;       TELEFAX:  (512) 499-6290
;   INFORMATION FOR SEQ ID NO:  71:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  644 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-911-894-71


alignment scores:
             Quality:     6.00              Length:       6
               Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2_COPY_21_29 x US-08-911-894-71    ..

 Align seq 1/1  to: US-08-911-894-71  from: 1  to: 644

        3 ValSerArgIleGlyLeu 8
          ||||||||||||||||||
     436 GTCAGCAGGATCGGCCTA 453

seq_name: /cgn2_6/ptodata/2/iia/6A_COMP_seq_US_08_911_894_71

seq_documentation_block:
; Sequence 5, Application US/08602093
; Patent No. 5847545
;  GENERAL INFORMATION:
;     APPLICANT:  Joseph, Rajiv
;     APPLICANT:  Dou, Dexian
;     TITLE OF INVENTION:  A NOVEL NEURONAL NEONATAL GENE
;     TITLE OF INVENTION:  NEURONATIN
;     NUMBER OF SEQUENCES:  24
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Kohn & Associates
;       STREET:  30500 No. 5847545thwestern Hwy.
;       CITY:  Farmington Hills
;       STATE:  Michigan
;       COUNTRY:  US
;       ZIP:  48334
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/602,093
;       FILING DATE:
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Kohn, Kenneth I.
;       REGISTRATION NUMBER:  30,955
;       REFERENCE/DOCKET NUMBER:  1059.00015
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (810) 539-5050
;       TELEFAX:  (810) 539-5055
;   INFORMATION FOR SEQ ID NO:  5:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  764 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-602-093-5


alignment scores:
             Quality:     6.00              Length:       6
               Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2_COPY_21_29 x US-08-602-093-5    ..

 Align seq 1/1  to: US-08-602-093-5  from: 1  to: 704

        2 ThrValSerArgThrGly 7
          ||||||||||||||||||
      411 ACGGTGTCGGCGGACGGG 428

seq_name: /cgn_6/ptodata/2/iia/6A_COMP_seq_US_08_602_093_2

seq_documentation_block:
; Sequence 2, Application US/08602093
; Patent No. 5847545
;  GENERAL INFORMATION:
;     APPLICANT:  Joseph, Rajiv
;     APPLICANT:  Dou, Dexian
;     TITLE OF INVENTION:  A NOVEL NEURONAL NEONATAL GENE
;     TITLE OF INVENTION:  NEURONATIN
;     NUMBER OF SEQUENCES:  24
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Kohn & Associates
;       STREET:  30500 No. 5847545thwestern Hwy.
;       CITY:  Farmington Hills
;       STATE:  Michigan
;       COUNTRY:  US
;       ZIP:  48334
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
```

```
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/602,093
;       FILING DATE:
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Kohn, Kenneth I.
;       REGISTRATION NUMBER:  30,995
;       REFERENCE/DOCKET NUMBER:  1059.00015
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (810) 539-5050
;       TELEFAX:  (810) 539-5055
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1113 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-602-093-2


Alignment_scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment_block:
 US-09-126-945B-2_COPY_21_29 x US-08-602-093-2    ..

 Align seq 1/1  to: US-08-602-093-2  from: 1  to: 1113

        2 ThrValSerArgThrGly 7
          ||||||||||||||||||
      146 ACGGTGTCCCGGACCGGG 163

seq_name: /cgn2_6/ptodata/2/iia/5B_COMB.seq:US-08-602-093-2

seq_documentation_block:
; Sequence 1, Application US/08602093
; Patent No. 5837535
;  GENERAL INFORMATION:
;     APPLICANT:  Joseph, Rajiv
;     APPLICANT:  Dou, Dexian
;     TITLE OF INVENTION:  A NOVEL NEURONAL NEONATAL GENE:
;     TITLE OF INVENTION:  NEURONATIN
;     NUMBER OF SEQUENCES:  23
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Kohn & Associates
;       STREET:  30500 No. 5837535thwestern Hwy.
;       CITY:  Farmington Hills
;       STATE:  Michigan
;       COUNTRY:  US
;       ZIP:  48334
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/602,093
;       FILING DATE:
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Kohn, Kenneth I.
;       REGISTRATION NUMBER:  30,995
;       REFERENCE/DOCKET NUMBER:  1059.00015
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (810) 539-5050
;       TELEFAX:  (810) 539-5055
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1195 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-602-093-1


Alignment_scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment_block:
 US-09-126-945B-2_COPY_21_29 x US-08-602-093-1    ..

 Align seq 1/1  to: US-08-602-093-1  from: 1  to: 1195

        2 ThrValSerArgThrGly 7
          ||||||||||||||||||
      227 ACGGTGTCCCGGACCGGG 244

seq_name: /cgn2_6/ptodata/2/iia/5B_COMB.seq:US-09-191-608-21

seq_documentation_block:
; Sequence 21, Application US/09191608
; Patent No. 6242216
; GENERAL INFORMATION:
;  APPLICANT: Lynch, Kevin J.
;  APPLICANT: Burnard, Edward C.
;  APPLICANT: Metzger, Randy E.
;  APPLICANT: Nikolatos, Wende
;  APPLICANT: Touma, Edward B.
;  APPLICANT: Van Biesen, T.
;  TITLE OF INVENTION: Nucleic Acids Encoding a Functional
;  TITLE OF INVENTION: Human Purinoreceptor P2X2 and P2X4 And Methods of Production
;  TITLE OF INVENTION: And Use Thereof
;  FILE REFERENCE: 6394.US.P1
;  CURRENT APPLICATION NUMBER: US/09/191,608
;  CURRENT FILING DATE: 1998-11-13
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 21
;   LENGTH: 1206
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-191-608-21


Alignment_scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment_block:
 US-09-126-945B-2_COPY_21_29 x US-09-191-608-21    ..

 Align seq 1/1  to: US-09-191-608-21  from: 1  to: 1206

        1 AspThrValSerArgThr 6
          ||||||||||||||||||
      739 GACACGGTTCCAGGACA 756

seq_name: /cgn2_6/ptodata/2/iia/5B_COMB.seq:US-08-742-621-2

seq_documentation_block:
; Sequence 2, Application US/08742621
; Patent No. 5856129
;  GENERAL INFORMATION:
;     APPLICANT:  HILLMAN, JENNIFER L.
;     APPLICANT:  CORLEMAN, ROGER
;     TITLE OF INVENTION:  NOVEL HUMAN PURINOCEPTOR
;     NUMBER OF SEQUENCES:  5
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Incyte Pharmaceuticals, Inc.
```

```
;       STREET:  4174 Porter Drive
;       CITY:  Palo Alto
;       STATE:  CA
;       COUNTRY:  US
;       ZIP:  94304
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ Version 1.5
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/234,529
;       FILING DATE:  Filed Herewith
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Billings, Lucy J.
;       REGISTRATION NUMBER:  36,749
;       REFERENCE/DOCKET NUMBER:  PF 0147 US
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415-855-0555
;       TELEFAX:  415 845-4166
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1762 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     IMMEDIATE SOURCE:
;       LIBRARY:
;       CLONE:  CONSENSUS
US-08-742-621-2


Alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-126-945B-2_COPY_21_29 x US-08-742-621-2  ..

 Align seq 1/1  to reverse of: US-08-742-621-2  from: 1  to: 1762

        1 AspThrValSerArgThr 6
          ||||||||||||||||||
     746 GACACAGTTCCAGGACA 763

Seq name: /seqn2_6/ptodata/2/ina/5A_COMP_seq/US-08-554-612C-13

Seq documentation block:
; Sequence 13, Application US/08554612C
; Patent No. 5747660
;  GENERAL INFORMATION:
;     APPLICANT:  Ozlicky, David
;     TITLE OF INVENTION:  PROSTAGLANDIN E2' RECEPTOR REGULATORY
;     TITLE OF INVENTION:  PROTEIN AND THERAPEUTIC USES
;     NUMBER OF SEQUENCES:  51
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Fish & Richardson P.C.
;       STREET:  2200 Sand Hill Road, Suite 100
;       CITY:  Menlo Park
;       STATE:  California
;       COUNTRY:  U.S.A.
;       ZIP:  94025
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/554,612C
;       FILING DATE:  No. 5747660ember 6, 1995
;       CLASSIFICATION:  536
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Sherwood, Pamela
;       REGISTRATION NUMBER:  36,677
;       REFERENCE/DOCKET NUMBER:  06519/004001
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 322 5070
;       TELEFAX:  (415) 854 0875
;   INFORMATION FOR SEQ ID NO:  14:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1784 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-554-612C-14


Alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-126-945B-2_COPY_21_29 x US-08-554-612C-14_rev  ..

 Align seq 1/1  to reverse of: US-08-554-612C-14  from: 1  to: 1784

        1 AspThrValSerArgThr 6
          ||||||||||||||||||
     1345 GACACAGTAAGTCCGACA 1328

Seq name: /seqn2_6/ptodata/2/ina/5A_COMP_seq/US-09-234-529-2

Seq documentation block:
; Sequence 2, Application US/09234529
; Patent No. 6096308
;  GENERAL INFORMATION:
;     APPLICANT:  Lal, Preeti
;     APPLICANT:  Hillman, Jennifer L.
;     APPLICANT:  Bandman, Olga
;     APPLICANT:  Corley, Neil C.
;     APPLICANT:  Shah, Purvi
;     TITLE OF INVENTION:  HUMAN PROTEIN KINASE AND KINASE INHIBITORS
;     NUMBER OF SEQUENCES:  9
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;       STREET:  3174 Porter Dr.
;       CITY:  Palo Alto
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  94304
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  09/09/234,529
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/972,816
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Billings, Lucy J.
;       REGISTRATION NUMBER:  36,749
;       REFERENCE/DOCKET NUMBER:  PF 0429 US
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  650 855 0555
;       TELEFAX:  650 845 4166
```

```
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1977 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  KIDNNOT25
;      CLONE:  3453694
US-09-231-529-2


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_21_29 x US-09-231-529-2/rev   ..

 Align seq 1/1  to reverse of: US-09-231-529-2  from: 1  to: 1977

        2 ThrValSerArgThrGly 7
          ||||||||||||||||||
       38 ACAGTGAGCAGGACTGGC 21

seq_name: [illegible] to reverse of: US-08-977-816-2

seq_documentation_block:
; Sequence 2, Application US/08977816
; Patent No. 6194186
;  GENERAL INFORMATION:
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Shah, Purvi
;    TITLE OF INVENTION:  HUMAN PROTEIN KINASE AND KINASE INHIBITORS
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  4174 Porter Dr.
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/977,816
;      FILING DATE:  Filed Herewith
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0429 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-855-0555
;      TELEFAX:  650-845-4166
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1977 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  KIDNNOT25
;      CLONE:  3453694
US-08-977-816-2


alignment scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_21_29 x US-08-977-816-2/rev   ..

 Align seq 1/1  to reverse of: US-08-977-816-2  from: 1  to: 1977

        2 ThrValSerArgThrGly 7
          ||||||||||||||||||
       38 ACAGTGAGCAGGACTGGC 21
```

```
M ut: US-09-126-945B-2_COPY_21_29 to: EST:*    -T_format : pir

Date: Sep 14, 2001  7:47 PM

About: Results were produced by the GenCore software, version 4.5.
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL framer_p2n.model -DEV xlp
 -Q "/cgn2_2/USERS/sp_c/258/5/2645/runat_11/US-09-126-945B-2_COPY_21_29/app_query.fasta_1.267
 -DB ESTN -QFMT fastap -SUFFIX olip2n.rst -QGAPOP 4.5000
 -GAPEXT 0.050 -MINMATCH 0.100 -TXGAPOP 0.000 -TXGAPEXT 0.000
 -QGAPOP 4.500 -QGAPEXT 0.050 -XGAPOP 60.000 -XGAPEXT 60.000
 -FGAPOP 6.000 -FGAPEXT 7.000 -YGAPOP 60.000 -YGAPEXT 60.000
 -DELOP 6.000 -DELEXT 7.000 -START 1 -MATRIX oligo
 -TRANS human40.cdi -LIST 45 -DOALIGN 200 -IBR_SCORE quality
 -THR_MIN 1 -ALIGN 15 -MODE local -OUTFMT pir -NORM cal -MINLEN 0
 -MAXLEN 2000000000 -USER US09126945_2CGN1_1_4645 -NCPU=6
 -ICPU 3 -LONGLOG -NO_XLPXY -WAIT -THREADS 1

Search Information block:
Query: US-09-126-945B-2_COPY_21_29
Query length: 9
Database: EST:*
Database sequences: 10228115
Database length: 431459454
Search time (sec): 9087.630000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000


score_list:
Sequence              Strd  Orig    ZScore    EScore  Len   ! Documentation ..
qb_est23:AI685592      +     9.00   168.20      2.99    486  ! AI685592 tw20c07.r1 NCI_CGAP_Pr
qb_est76:BE563288      +     9.00   162.77      6.01    938  ! BE563288 601345410F1 NIH_MGC_34
qb_est7:AA460196       +     7.00   134.47    226.43    249  ! AA460196 zx50d06.r1 Soares_test
qb_est65:BB451783      +     7.00   133.35    261.65    299  ! BB451783 BB451783 RIKEN full-le
qb_est7:AA466691       -     7.00   133.26    264.11    303  ! AA466691 zd95c06.r1 Soares mous
qb_est94:BF959653      -     7.00   132.48    292.29    341  ! BF959653 RC5 NN0249 041208 031
qb_qss8:AQ213391       -     7.00   132.24    301.64    358  ! AQ213391 HS_4048_B1_F02_MR CIT
qb_qss28:AZ198525      +     7.00   132.18    303.64    361  ! AZ198525 IMAGE:[illegible] M
qb_est25:AI814300      +     7.00   132.00    310.92    372  ! AI814300 QI M:ALL adm 3 09 C QI
qb_est25:AI852126      +     7.00   132.00    310.92    372  ! AI852126 QI M:BB0-ajb : 07 0 QI
qb_est7:AA414908       +     7.00   131.74    321.44    388  ! AA414908 ve59e02.r1 Knowles Sol
qb_est85:BF401985      -     7.00   131.63    326.01    395  ! BF401985 602032823F1 NCI_CGAP_
qb_est10:AA646533      -     7.00   131.45    333.81    407  ! AA646533 wq76g02.s1 Knowles Sol
qb_est11o:W71676       -     7.00   131.39    336.40    411  ! W71676 mc84h02.r1 Soares mouse
qb_est77:BE626836      +     7.00   131.40    340.27    417  ! BE626836 ut85a04.x1 Soares_mamm
qb_est14:AI006006      -     7.00   131.09    349.27    431  ! AI006006 qa84q08.r1 Soares_mamm
qb_est81:BE954803      -     7.00   131.09    349.27    431  ! BE954803 QI-M-CF1-bri b-11-0-QI
qb_qss9:AQ264771       +     7.00   131.06    350.55    433  ! AQ264771 CITBI E1 2510M18.TF C
qb_est97:BG143540      +     7.00   130.90    358.20    445  ! BG143540 qt56q10.y1 Soares_mesa
qb_qss9:AQ270667       +     7.00   130.90    358.20    445  ! AQ270667 HS_2010_B2_F02_17 CIT
qb_est20:AI426325      +     7.00   130.84    360.71    449  ! AI426325 mq58d11.x1 Soares_mous
qb_est96:BG062189      +     7.00   130.80    362.64    452  ! BG062189 [illegible] Mouse Newb
qb_est80:BE854360      -     7.00   130.79    363.27    454  ! BE854360 nx29d04.y1 Soares_NMMA
qb_est77:BE628576      -     7.00   130.75    365.17    456  ! BE628576 ut04b03.x1 Soares_thy
qb_est98:BG244061      -     7.00   130.75    365.17    456  ! BG244061 602359124F1 NCI_CGAP_M
qb_est90:BF660374      +     7.00   130.63    370.86    465  ! BF660374 ma05d09.x1 Soares_thy
qb_est110:W83305       +     7.00   130.57    373.37    469  ! W83305 zb64b04.r1 Soares_fetal_
qb_est10:AA681340      -     7.00   130.48    377.27    476  ! AA681340 vr77eu7.s1 Knowles Sol
qb_est48:AW539918      +     7.00   130.11    396.45    506  ! AW539918 C0126B06 3 Mouse E7.5
qb_est23:AI658220      -     7.00   130.01    401.40    514  ! AI658220 tv20c06.y1 Zebrafish W
qb_qss29:AZ594978      +     7.00   129.88    408.17    525  ! AZ594978 1M0407117F Mouse 10kb
qb_est96:BG064433      +     7.00   129.80    412.46    532  ! BG064433 H3019F03 3 NIA Mouse 1
qb_est48:AW555740      -     7.00   129.77    413.68    534  ! AW555740 L0258C10 3 Mouse Newbo
qb_qss20:AZ061960      -     7.00   129.72    416.71    539  ! AZ061960 EPT1 2 415N23.TJ EPT1
qb_est84:BF178248      -     7.00   129.65    420.40    545  ! BF178248 601867779F1 NCI_CGAP_M
qb_est48:AW544652      -     7.00   129.49    428.91    559  ! AW544652 C0184C04 3 Mouse E7.5
qb_qss25:AZ368189      -     7.00   129.28    440.99    579  ! AZ368189 1M3119H12F Mouse 10kb
qb_est48:AW542649      +     7.00   129.24    442.79    582  ! AW542649 C0160B04 3 Mouse E7.5
qb_est73:BE871008      -     7.00   129.13    449.29    593  ! BE871008 601219608F1 NCI_CGAP_

qb_est23:AI685592      7.00  129.02  461.32    618  ! AI685592 tw20c07.r1 NCI_CGAP_Pr
qb_qss5:FK002608       -     7.00  128.86  464.88    619  ! AL025977 Fugu rubripes GSS s
qb_est54:AW507116      -     7.00  128.85  465.47    620  ! AW507116 QV3-BN0047 1504-c1
qb_est100:BG414849     -     7.00  128.66  477.29    640  ! BG414849 RVSMEK00451 Mixed
qb_est47:AW451117      +     7.00  128.48  488.45    659  ! AW451117 UI-R-BI5-ajq-g-03-0
qb_qss25:AZ367120      7.00  128.42  491.96    665  ! AZ367120 IMAGE:[illegible] M

Seq name: qb_est23:AI685592

Seq documentation block:
LOCUS       AI685592          486 bp    mRNA            EST       15-FEB-[illegible]
DEFINITION  tw20c07.r1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2251596 3'
            similar to contains PIR5.13 MSR1 repetitive element ;, mRNA
            sequence.
ACCESSION   AI685592
VERSION     AI685592.1  GI:4896886
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 486)
  AUTHORS   NCI [illegible] http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www.bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1143  Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 480.
FEATURES             Location/Qualifiers
     source          1..486
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2251596"
                     /clone_lib="NCI_CGAP_Pr28"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="organ: prostate; Vector: pT7T3-Pac (Pharmacia)
                     with a modified polylinker; Plasmid DNA from the
                     normalized library NCI_CGAP_Pr22 was prepared, and ss
                     circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR amplified cDNAs from a pool
                     of 5,000 clones made from the same library (colonies
                     [illegible], 11011w2-11011959, and 12179z8-12200i5).
                     Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT        72 a     146 c     110 g      64 t
ORIGIN


alignment scores:
            Quality:     9.00              Length:      9
              Ratio:    1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_21_29 x AI685592  ..

 Align seq 1/1  to: AI685592  from: 1  to: 486

        1 AspThrValSerArgThrGlyLeuGlu 9
          ||||||||||||||||||||||||||||
```

```
     221 GACACGGTGTCGCGGACAGGCTTGGAG 247

seq_name: qb_est76:BE563288

seq_documentation_block:
LOCUS       BE563288      933 bp    mRNA            EST       15 AUG 2000
DEFINITION  601339410F1 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:3689402 5',
            mRNA sequence.
ACCESSION   BE563288
VERSION     BE563288.1  GI:9807100
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 933)
  AUTHORS   NIH_MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Lisa Hong/Robin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM384  row: c  column: 03
            High quality sequence stop: 678.
FEATURES             Location/Qualifiers
     source          1..933
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3689402"
                     /clone_lib="NIH_MGC_99"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed
                     by Ling Hong in the laboratory of Gerald M. Rubin
                     (University of California, Berkeley) using ZAP-cDNA
                     synthesis kit (Stratagene) and Superscript II RI (Life
                     Technologies)."
BASE COUNT       148 a    316 c    319 g    150 t
ORIGIN


alignment scores:
            Quality:     9.00              Length:       9
              Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2 (COPY_21_29 x BE563288    ..

 Align seq 1/1  to: BE563288  from: 1  to: 933

        1 AspThrValSerArgThrGlyLeuGlu 9
          |||||||||||||||||||||||||||
      455 GACACGGTGTCGCGGACAGGCTTGGAG 481

seq_name: qb_est7:AA460196

seq_documentation_block:
LOCUS       AA460196      249 bp    mRNA            EST       09 JUN 1997
DEFINITION  zx50d06.r1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:795659
            5', mRNA sequence.
ACCESSION   AA460196
VERSION     AA460196.1  GI:2185012
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 249)
  AUTHORS   Keller,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Morris,M., Martin,J., Moore,B.
            , Schellenberg,K., Steptoe,M., Ian,F., Theising,B., White,Y., Wylie
            ,T., Waterston,R. and Wilson,R.
  TITLE     Wash0 Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Putative full length read
            The vector to vector length is 250
            Seq primer:  28mlB rev2 ET from Amersham.
FEATURES             Location/Qualifiers
     source          1..249
                     /organism="Homo sapiens"
                     /db_xref="GDB:6043904"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:795659"
                     /clone_lib="Soares_testis_NHT"
                     /sex="male"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech Laboratories
                     , Inc., and primed with  a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT        83 a     40 c     53 g     73 t
ORIGIN


alignment scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2 (COPY_21_29 x AA460196    ..

 Align seq 1/1  to: AA460196  from: 1  to: 249

        3 ValSerArgThrGlyLeuGlu 9
          |||| ||||||||||||||||
      138 GtgAgtAGGAcAGGCTTGGAG 158

seq_name: qb_est65:BB451783

seq_documentation_block:
LOCUS       BB451783      299 bp    mRNA            EST       21 JUL 2000
DEFINITION  BB451783 RIKEN full length enriched, 12 days embryo spinal ganglion
            Mus musculus cDNA clone D130012L20 3' similar to AF166382 Mus
            musculus serpentine receptor (Cyt28) mRNA, mRNA sequence.
ACCESSION   BB451783
VERSION     BB451783.1  GI:9412818
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
```

REFERENCE 1 (bases 1 to 299)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,I., Carninci
,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,I., Itoh,M.,
Izawa,M., Kadota,K., Kamiwa,I., Kato,S., Kawai,J., Kibayashi,N.,
Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
, Matsuyama,T., Miki,E., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
, Ono,I., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
,Y., Shigemoto,Y., Shinagawa,A., Shiraki,I., Sogabe,Y., Suwahara,Y.
, Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
, Yano,R., Yasunishi,A., Yokota,I., Yoshida,K., Yoshiki,A., Yoshino
,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL Unpublished (2000)
COMMENT Contact: Yoshihide Hayashizaki
Genome Exploration Research Group, Life Science Tsukuba Center,
Genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic
Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: +81-298-36-9013
Fax: +81-298-36-9098
Email: genome-res@rtc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 464-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
FEATURES Location/Qualifiers
source 1..299
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="D1400121L20"
/clone_lib="RIKEN full-length enriched, 12 days embryo
spinal ganglion"
/tissue_type="spinal ganglion"
/dev_stage="12 days embryo"
/lab_host="DH10B"
/note="Site_1: Sall; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in RIKEN,
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTVN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGAGATTCTCGAGTTAATTAAATTAATGCGGCCGCNNN 3']. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT 79 a 87 c 46 g 87 t
ORIGIN

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_21_29 x BB451783 ..

Align set 1/1 to: BB451783 from: 1 to: 299

5 ValSerArgIleGlyLeuGln 9
|||.|.|||||||||||||||||
91 GTTTCACGCACGGGCCTCGAA 111

Seq_name: gb_est7:AA466691

seq_documentation_block:
LOCUS AA466691 404 bp mRNA EST 11-JUN-1997
DEFINITION vd95e06.r1 Soares mouse NbMH Mus musculus cDNA clone IMAGE:808454
5', mRNA sequence.
ACCESSION AA466691
VERSION AA466691.1 GI:2192881
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 404)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Bystrom,N., Dubuque,T.,
Gelfand,C., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Stoptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Pennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:484098
Seq primer: -28m13 revz ET from Amersham
High quality sequence stop: 234.
FEATURES Location/Qualifiers
source 1..404
/organism "Mus musculus"
/strain "C57BL/6J"
/db_xref "taxon:10090"
/clone "IMAGE:808454"
/clone_lib "Soares mouse NbMH"
/sex "male"
/tissue_type "heart"
/dev_stage "4 weeks"
/lab_host="DH10B"
/note "Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGGAGCGGCCGCGAAAGTTTTTTTTTTTTTTTTT
3']; double stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT7T3 vector. RNA
provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."
BASE COUNT 49 a 94 c 80 g 80 t
ORIGIN

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```
alignment block:
 US-09-126-945B-2_COPY_21_29 x AA466691/rev   ..

 Align seq 1/1  to reverse of: AA466691  from: 1  to: 344

        1 AspThrValSerArgThrGly 7
          |||||||||||||||||||||
      205 GACACTGTGTCAAGGACAGGG 185

seq_name: gb:est94:BF959653

seq_documentation_block:
LOCUS        BF959653        344 bp    mRNA            EST       22-JAN-2001
DEFINITION   RC5-NN0243-041209-C41-D02-NN0243 Homo sapiens cDNA, mRNA sequence.
ACCESSION    BF959653
VERSION      BF959653.1  GI:12376928
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 344)
  AUTHORS    Dias Neto,E., Garcia Correa,R., Verjovski Almeida,S., Briones,M.R.,
             Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
             Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
             Brumstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
             ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
             Simpson,A.J.
  TITLE      Shotgun sequencing of the human transcriptome with ORF expressed
             sequence tags
  JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE    20202663
COMMENT      Contact: Simpson A.J.G.
             Laboratory of Cancer Genetics
             Ludwig Institute for Cancer Research
             Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo SP,
             Brazil
             Tel: +55 11 2704922
             Fax: +55-11-2707001
             Email: asimpson@ludwig.org.br
             This sequence was derived from the EARESP/LICR Human Cancer Genome
             Project. This entry can be seen in the following URL
             (http://www.ludwig.org.br/cgi-bin/getbest2.pl.tt?"RC5-NN0243-
             041200-031-D02&t3-2000-12-04&t4-1)
             Seq primer: puc 18 forward
             High quality sequence start: 20
             High quality sequence stop: 344.
FEATURES             Location/Qualifiers
     source          1..344
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="NN0243"
                     /dev_stage="Adult"
                     /note="organ: nervous_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       115 a     74 c     63 g     92 t
ORIGIN


alignment scores:
            Quality:    7.00            Length:      7
              Ratio:   1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_21_29 x BF959653/rev

 Align seq 1/1  to reverse of: BF959653  from: 1  to: 344
```

```
        1 AspThrValSerArgThrGly 7
          |||||||||||||||||||||
      145 GACACTGTGTCAAGGACAGGG 125

seq_name: gb:est98:AQ213391

seq_documentation_block:
LOCUS        AQ213391        358 bp    DNA             GSS       18-SEP-1998
DEFINITION   HS_3048_B1_F02_MR_CIT Approved Human Genomic Sperm Library D Homo
             sapiens genomic clone Plate=3048 Col=3 Row=L, DNA sequence.
ACCESSION    AQ213391
VERSION      AQ213391.1  GI:3624592
KEYWORDS     GSS.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 358)
  AUTHORS    Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
             Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
             Hood,L.
  TITLE      Sequence tagged connectors: A sequence approach to mapping and
             scanning the human genome
  JOURNAL    Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE    99380589
COMMENT      Contact: Mahairas GG, Wallace JC, Hood L
             High Throughput Sequencing Center
             University of Washington
             401 Queen Anne Avenue North, Seattle, WA 98109, USA
             Tel: (206) 616-3618
             Fax: (206) 616-3887
             Email: jwallace@u.washington.edu
             Sequence Tagged Connector
             Plate: 3048  row: L  column: 3
             Class: BAC ends
             High quality sequence stop: 358.
FEATURES             Location/Qualifiers
     source          1..358
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate:3048 Col:3 Row:L"
                     /clone_lib:"CIT Approved Human Genomic Sperm Library D"
                     /sex:"male"
                     /note "Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E Coli DH10B"
BASE COUNT        76 a    114 c     80 g     88 t
ORIGIN


alignment scores:
            Quality:    7.00            Length:      7
              Ratio:   1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_21_29 x AQ213391/rev   ..

 Align seq 1/1  to reverse of: AQ213391  from: 1  to: 358

        2 ThrValSerArgThrGlyLeu 8
          |||||||||||||||||||||
      338 ACGGTCTCAAGGACAGGGTTG 318

seq_name: gb:gss28:AZ498525

seq_documentation_block:
LOCUS        AZ498525        361 bp    DNA             GSS       05-OCT-2000
DEFINITION   IM0435D19R Mouse 10kb plasmid RUGCIM library Mus musculus genomic
             clone RUGCIM0435D19 R, DNA sequence.
ACCESSION    AZ498525
VERSION      AZ498525.1  GI:10676457
KEYWORDS     GSS.
```

```
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 361)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beaucorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,P., Stokes,R., Tinney,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., 2123, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000  Std Error: 0.00
            Plate: 0435  row: D  column: 19
            Seq primer: CACACAGGAAACAGCTATGAC
            Class: plasmid ends
            High quality sequence stop: 361.
FEATURES             Location/Qualifiers
     source          1..361
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0435D19"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, Tet resistant, F'"
                     /note="Vector: PWD42mv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/cmr/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at a constant pressure. The 3' overhang
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high molar excess. The
                     adaptored DNA was purified and size selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      100 a      64 c      85 g     112 t
ORIGIN


alignment_scores:
            Quality:     7.00            Length:       7
              Ratio:    1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_21_29 x AZ498525  ..

 Align seq 1/1  to: AZ498525  from: 1  to: 361

       2 ThrValSerArgThrGlyLeu 8
         |||||||||||||||||||||
     284 ACAGTCTCIAGGACGGGACTG 303

seq_name: qb_est25:AI844300
```

```
seq_documentation_block:
LOCUS       AI844300        472 bp    mRNA            EST       15-JUL-1999
DEFINITION  UI M-ALI ahm d-09 0-01.s1 NIH_BMAP_MCO_N Mus musculus cDNA clone
            UI M-ALI ahm d-09 0-01 3', mRNA sequence.
ACCESSION   AI844300
VERSION     AI844300.1  GI:5188206
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 472)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9663, Bethesda, MD
            20892-9663, USA
            Tel: 301 443 1709
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the Not1 site
            and the oligo-dT track served to verify it as a clone from the
            normalized prefrontal cortex library cDNA Library Preparation: M.B.
            Soares lab Clone distribution: NIH BMAP cDNA clones will be made
            available by the means that is soon to be determined. When NIH
            determines the means for distribution of the BMAP cDNA clones, this
            record will be updated accordingly when that means is determined.
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..472
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI M ALI-ahm d-09 0 UI"
                     /clone_lib "NIH_BMAP_MCO_N"
                     /dev_stage "27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /note "Vector: pT73D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not 1; Site_2: Eco R1; The
                     NIH_BMAP_MCO_N library is a normalized library constructed
                     from  mouse cortex. The tag is a string of 5 nucleotides
                     present  between the Not 1 site and the oligo-dT track.
                     The library was  constructed as described by Bonaldo,
                     Lennon and Soares,  Genome Research 6: 791-806, 1996.
                     Tissue provided by Mr. Annie Novakovich, Vicki Miller
                     Laboratories.
                     TAG_LIB NIH_BMAP_MCO_N
                     TAG_TISSUE prefrontal-cortex
                     TAG_SEQ GGGCA"
BASE COUNT      116 a      89 c      85 g      82 t
ORIGIN


alignment_scores:
            Quality:     7.00            Length:       7
              Ratio:    1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_21_29 x AI844300  ..

 Align seq 1/1  to: AI844300  from: 1  to: 472

       1 AspThrValSerArgThrGly 7
         |||||||||||||||||||
```

```
      330 GACACTGTGTTCAAGGACAGGG 450

seq name: qb_est2:AI852126

seq documentation block:
LOCUS       AI852126       472 bp    mRNA             EST       15 JUL 1999
DEFINITION  UI-M-BH0-ajb-c-07-0-UI.s1 NIH_BMAP_M_S1 Mus musculus cDNA clone
            UI-M-BH0-ajb-c-07-0-UI 3', mRNA sequence.
ACCESSION   AI852126
VERSION     AI852126.1  GI:5496032
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 472)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N 7190, MSC 9643, Bethesda, MD
            20892 9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            normalized prefrontal cortex library cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: NIH BMAP cDNA clones will be made
            available by the means that is soon to be determined. When NIH
            determines the means for distribution of the BMAP cDNA clones, this
            record will be updated accordingly when that means is determined.
            Seq primer: M13 Forward
            POLYA Yes.
FEATURES             Location/Qualifiers
     source          1..472
                     /organism "Mus musculus"
                     /strain "C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone "UI-M-BH0-ajb-c-07-0-UI"
                     /clone_lib "NIH_BMAP_M_S1"
                     /dev_stage "27-42 days"
                     /lab_host "DH10B (Life Technologies)"
                     /note "Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_M_S1 library is a subtracted library derived from
                     a mixture of normalized libraries from ten regions of the
                     mouse brain (cerebellum, brain stems, olfactory bulbs,
                     hypothalamus, cortex, amygdala, basal ganglia, pineal
                     gland, striatum, hippocampus). The driver used for
                     subtraction consisted of a pool of 20,000 cDNA clones
                     obtained from non normalized and normalized libraries of
                     these ten regions of the mouse brain.
                     TAG_LIB NIH_BMAP_M_S1
                     TAG_TISSUE prefrontal cortex
                     TAG_SEQ GCTCA"
BASE COUNT      116 a      90 c      84 g      82 t
ORIGIN


alignment scores:
            Quality:     7.00            Length:       7
              Ratio:    1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_21_29 x AI852126  ..

 Align seq 1/2  to: AI852126  from: 1  to: 472

        1 AspThrValSerArgThrGly 7
          ||||||||||||||||||||
      329 GACACTGTGTTCAAGGACAGGG 349

seq name: qb_est2:AA414908

seq documentation block:
LOCUS       AA414908       388 bp    mRNA             EST       08 MAY 1997
DEFINITION  vc50c02.r1 Knowles Solter mouse 2-cell Mus musculus cDNA clone
            IMAGE:777986 3', mRNA sequence.
ACCESSION   AA414908
VERSION     AA414908.1  GI:2075148
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 388)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU HHMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:470842.
FEATURES             Location/Qualifiers
     source          1..388
                     /organism "Mus musculus"
                     /strain "B6D2 F1/J"
                     /db_xref "taxon:10090"
                     /clone "IMAGE:777986"
                     /cell_line "Knowles Solter mouse 2-cell"
                     /tissue_type="embryo"
                     /dev_stage "2-cell"
                     /lab_host "DH10B"
                     /note "organ: embryo; Vector: pBluescribe (modified);
                     Site_1: MluI; Site_2: SalI; Cloned unidirectionally from
                     mRNA prepared from 13,500 2-cell stage embryos.  Primer:
                     SalI(dT): 5' CGGTCGACCGTCGACCGTTTTTTTTTTTTT 3'.  cDNAs
                     were cloned into the MluI/SalI sites of a modified
                     pBluescribe  vector using commercial linkers (NEB).
                     Average insert size:  1.2 kb."
BASE COUNT      124 a      94 c      99 g      71 t
ORIGIN


alignment scores:
            Quality:     7.00            Length:       7
              Ratio:    1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_21_29 x AA414908  ..

 Align seq 1/2  to: AA414908  from: 1  to: 388

        1 AspThrValSerArgThrGly 7
          ||||||||||||||||||||
      316 GACACTGTGTTCAAGGACAGGG 336
```

```
Seq_name: gb_est85:BF301985

Seq_documentation_block:
LOCUS       BF301985       395 bp    mRNA            EST       21 NOV 2000
DEFINITION  yo20a02s1 NCI_CGAP_Sc2 Mus musculus cDNA clone IMAGE:4168108 5'
            mRNA sequence.
ACCESSION   BF301985
VERSION     BF301985.1  GI:11248508
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 395)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9459  row: m  column: 05
            High quality sequence stop: 493.
FEATURES             Location/Qualifiers
     source          1..395
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4168108"
                     /clone_lib="NCI_CGAP_Sc2"
                     /lab_host="DH10B (11 phage resistant)"
                     /note="Organ: salivary gland; Vector: pCMV SPORT6; Site_1:
                     NotI; Site_2: SalI; Cloned unidirectionally. Primer: oligo
                     dT. Average insert size 1.3 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT       86 a     93 c     99 g    117 t
ORIGIN


alignment_scores:
            Quality:     7.00            Length:       7
              Ratio:    1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_21_29 x BF301985/rev   ..

 Align seq 1/1  to reverse of: BF301985  from: 1  to: 395

       1 AspThrValSerArgThrGly 7
         |||||||||||||||||||||
      70 GACACTGTGTCAAGGACAGGG 50

Seq_name: gb_est10:AA636533

Seq_documentation_block:
LOCUS       AA636533       407 bp    mRNA            EST       22-OCT-1997
DEFINITION  vq76g02.s1 Knowles Solter mouse 2-cell Mus musculus cDNA clone
            IMAGE:1108274 5', mRNA sequence.
ACCESSION   AA636533
VERSION     AA636533.1  GI:2560312
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 407)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Stoptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:606442
            High quality sequence stop: 444.
FEATURES             Location/Qualifiers
     source          1..407
                     /organism="Mus musculus"
                     /strain="B6SJL F1/J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1108274"
                     /clone_lib="Knowles Solter mouse 2-cell"
                     /tissue_type="embryo"
                     /dev_stage="2-cell"
                     /lab_host="DH10B"
                     /note="Organ: embryo; Vector: pBluescribe (modified);
                     Site_1: MluI; Site_2: SalI; Cloned unidirectionally from
                     mRNA prepared from 13,500 2-cell stage embryos. Primer:
                     SalI(dT): 5'-CGGTCGACCGTCGACGGTTTTTTTTTTTTTT-3'.  cDNAs
                     were cloned into the MluI/SalI sites of a modified
                     pBluescribe vector using commercial linkers (NEB).
                     Average insert size: 1.2 kb."
BASE COUNT       69 a    128 c    107 g    103 t
ORIGIN


alignment_scores:
            Quality:     7.00            Length:       7
              Ratio:    1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_21_29 x AA636533/rev   ..

 Align seq 1/1  to reverse of: AA636533  from: 1  to: 407

       1 AspThrValSerArgThrGly 7
         |||||||||||||||||||||
     265 GACACTGTGTCAAGGACAGGG 245

Seq_name: gb_est110:W71676

Seq_documentation_block:
LOCUS       W71676         411 bp    mRNA            EST       17-JUN-1996
DEFINITION  mo44h02.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:389427 5', mRNA sequence.
ACCESSION   W71676
VERSION     W71676.1  GI:1380417
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 411)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Stoptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
```

```
COMMENT     Contact: Marra M/Mouse EST Project
            WashU RIMM Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:241259
            Seq primer: ETPrimer
            High quality sequence stop: 442.
FEATURES             Location/Qualifiers
     source          1..411
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:489427"
                     /clone_lib="Soares mouse embryo, NbME13.5 14.5"
                     /sex="unknown"
                     /tissue_type="embryo"
                     /dev_stage="13.5-14.5dpe total fetus"
                     /lab_host="DH10B"
                     /note="Vector: pT7T4D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I   oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTTT
                     T 3'], on equal amounts of mRNA from 2 13.5dpe and 2
                     14.5dpe embryos [total RNA provided by  Minoru Ko, Wayne
                     State Univ., from 2 ]; double stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector. Library went through one round of
                     normalization, and was constructed by Bento Soares and
                     M.Fatima Bonaldo.  "
BASE COUNT        71 a    107 c    109 g    124 t
ORIGIN


alignment scores:
            Quality:    7.00            Length:      7
              Ratio:   1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_21_29 x W71676/rev    .

 Align seq 1/1  to reverse of: W71676  from: 1  to: 411

        1 AspThrValSerArgThrGly 7
          |||||||||||||||||||||
      118 GACACTGTGTCAAGGACAGGG 98

seq name: gb_est77:BE626836

seq documentation block:
LOCUS       BE626836       417 bp    mRNA            EST       24 AUG 2000
DEFINITION  ut85a04.x1 Soares mammary_gland NMLMG Mus musculus cDNA clone
            IMAGE:3369198 3', mRNA sequence.
ACCESSION   BE626836
VERSION     BE626836.1  GI:9907246
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 417)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Other ESTs: ut85a04.y1
            Contact: Robert Strausberg, Ph.D.
            Email: cgaps@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1078802
            High-quality sequence stop: 393.
FEATURES             Location/Qualifiers
     source          1..417
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3369198"
                     /clone_lib="Soares mammary gland NMLMG"
                     /sex="female (lactating)"
                     /tissue_type="mammary gland"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from mammary
                     gland tissue from a lactating female, and was then primed
                     with a Not I   oligo(dT) primer. Double stranded  cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I  and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector.  Library is normalized.  Library
                     was constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       130 a    104 c    107 g     77 t
ORIGIN


alignment scores:
            Quality:    7.00            Length:      7
              Ratio:   1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_21_29 x BE626836    ..

 Align seq 1/1  to: BE626836  from: 1  to: 417

        1 AspThrValSerArgThrGly 7
          |||| ||||||||||||||||
      417 GACACTGTGTCAAGGACAGGG 437
```

```
CM of: US-09-126-945B-2_COPY_38_46 to: GenEmbl:*   out_format : pts

Date: Sep 14, 2001  9:36 PM

About: Results were produced by the GenCore software, version 4.5.
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL frames_p2n.model -DEV xlp
 -Q "/cgn2_1/tmp/US-09-126-945B-2_COPY_38_46.seq" -DB GenEmbl -QFMT fastap -SUFFIX olip2n.rge -GAPOP 4.500
 -GAPEXT 0.050 -MINMATCH 0.100 -LOOPCL 0.000 -LOOPEXT 0.000
 -QGAPOP 4.500 -QGAPEXT 0.050 -XGAPOP 60.000 -XGAPEXT 60.000
 -FGAPOP 6.000 -FGAPEXT 7.000 -YGAPOP 60.000 -YGAPEXT 60.000
 -DELOP 6.000 -DELEXT 7.000 -START 1 -MATRIX oligo
 -TRANS human40.cdi -LIST 45 -DOCALIGN 200 -THR_SCORE quality
 -THR_MIN 1 -ALIGN 1% -MODE LOCAL -OUTFMT pts -NORM ext -MINGEN 6
 -MAXLEN 2000000000 -USER US09126945B_COPY_1_7594 -NPRO 4
 -INPU 4 -LONGLOG -NO_XLPXY -WAIT -THREADS 1

Search information block:
Query: US-09-126-945B-2_COPY_38_46
Query length: 9
Database: GenEmbl:*
Database sequences: 1344157
Database length: -856060004
Search time (sec): 6503.700000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000


score_list:
Sequence              Strd Orig    %Score    ES-ES  Len  ! Documentation
qb_pat1:AX048148        +     9.00  174.97   0.1649    1894  ! AX048148 Sequence 1 from Patent
qb_pr4:AF071538         +     9.00  174.97   0.1649    1894  ! AF071538 Homo sapiens Ets tran
qb_pr1:AB031549         +     9.00  174.95   0.1651    1898  ! AB031549 Homo sapiens ISE mRNA
qb_pat1:AX048159        +     9.00  171.82   0.2470    3317  ! AX048159 Sequence 12 from Pate
qb_pr6:AL157372         -     9.00  150.37   3.82      150660  ! AL157372 Human DNA sequence
qb_pr1:AC005192         -     8.00  130.72   48.05     128680  ! AC005192 Homo sapiens BAC cl
qb_pr7:BC002903         -     7.00  134.36   30.14     1743  ! BC002903 Homo sapiens, clone M
qb_ro2:RATENKA4         +     7.00  133.25   34.73     2122  ! M32784 Rat dynorphin gene, exo
qb_pr5:AK026996         +     7.00  131.72   42.21     2784  ! AK026996 Homo sapiens cDNA: FL
qb_ro2:RATENKB          +     7.00  131.37   44.18     2963  ! M10088 Rat prodynorphin (Prepr
qb_pr1:AB046773         +     7.00  127.86   69.35     5538  ! AB046773 Homo sapiens mRNA for
qb_ba3:RFCGCP           -     7.00  126.30   84.67     7404  ! X56082 R.flavefaciens DNA sequ
qb_bal:AE000078         -     7.00  122.04   146.22    15586  ! AE000078 Rhizobium sp. NGR234
qb_ba3:SCD35            +     7.00  117.28   269.47    36394  ! AL160412 Streptomyces coelico
qb_ba3:SCL1C2           -     7.00  116.44   299.86    42210  ! AL031124 Streptomyces coelico
qb_htq9:AC024308        -     7.00  114.35   392.33    61278  ! AC024308 Homo sapiens clone R
qb_pr8:HS1014D13        -     7.00  113.50   437.19    71292  ! AL022411 Human DNA sequence f
qb_htq8:AC022964        -     7.00  112.80   478.36    80684  ! AC022964 Homo sapiens chromos
qb_pr5:AL135784         -     7.00  112.32   528.69    92697  ! AL135784 Human DNA sequence f
qb_htq2:AC009984        -     7.00  111.46   568.30    102468  ! AC009984 Drosophila melanoga
qb_htq4:AC014977        -     7.00  111.24   584.66    106584  ! AC014977 Drosophila melanoga
qb_in1:AC008140         -     7.00  109.66   715.24    140973  ! AC008140 Drosophila melanoga
qb_ro1:AC009725         +     7.00  109.56   736.46    145152  ! AC009725 Mus musculus chromo
qb_htq7:AC021328        +     7.00  109.44   746.63    146856  ! AC021328 Homo sapiens clone
qb_htq8:AC022545        -     7.00  109.32   747.68    149919  ! AC022545 Homo sapiens clone
qb_htq7:AC021815        +     7.00  109.00   778.72    158623  ! AC021815 Homo sapiens chromo
qb_in1:AC009219         -     7.00  108.99   780.08    159007  ! AC009219 Drosophila melanoga
qb_htq23:AP001773       +     7.00  108.88   790.99    162102  ! AP001773 Homo sapiens chromo
qb_htq7:AC021331        -     7.00  108.79   800.41    164794  ! AC021331 Homo sapiens clone
qb_htq10:AC025705       +     7.00  108.61   819.15    170161  ! AC025705 Homo sapiens chromo
qb_htq21:AL359977       -     7.00  108.60   820.42    170530  ! AL359977 Homo sapiens chromo
qb_htq16:AC079115       +     7.00  108.29   853.61    180173  ! AC079115 Homo sapiens chromo
qb_htq5:AC016916        +     7.00  108.21   861.68    182540  ! AC016916 Homo sapiens chromo
qb_htq20:AL355586       +     7.00  108.11   872.88    185849  ! AL355586 Homo sapiens chromo
qb_pr2:AC006459         -     7.00  108.08   876.74    186981  ! AC006459 Homo sapiens BAC cl
qb_htq15:AC073389       +     7.00  107.99   886.45    189857  ! AC073389 Homo sapiens chromo
qb_htq23:AP000915       +     7.00  107.99   886.47    189864  ! AP000915 Homo sapiens chromo
qb_htq24:AP001853       -     7.00  107.96   890.10    190944  ! AP001853 Homo sapiens chromo
qb_htq24:AP001649       +     7.00  107.93   893.91    192078  ! AP001649 Homo sapiens chromo
qb_htq15:AC073544       -     7.00  107.64   927.91    202288  ! AC073544 Homo sapiens chr
qb_htq19:AC090966       -     7.00  107.58   934.76    204361  ! AC090966 Rapio cynocephal
qb_htq16:AC074306       -     7.00  107.22   979.25    217978  ! AC074306 Mus musculus chr
qb_htq4:AC015537        +     7.00  107.12   990.85    221569  ! AC015537 Homo sapiens chr
qb_in2:AE003688         -     7.00  107.04   1.0e+03   224795  ! AE003688 Drosophila melan
qb_pat1:A79350          -     7.00  105.06   1.3e+03   320040  ! A79350 S pence 1 from Pa

seq_name: qb_pat1:AX048148

seq_documentation_block:
LOCUS       AX048148      1894 bp    DNA             PAT       15-SEP-2000
DEFINITION  Sequence 1 from Patent WO0070092.
ACCESSION   AX048148
VERSION     AX048148.1  GI:11876967
KEYWORDS
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1894)
  AUTHORS   Kaufmann,J. and Xin,H.
  TITLE     Expression of ets domain proteins in cancer
  JOURNAL   Patent: WO 0070092-A 1 23-NOV 2000;
            Chiron Corporation (US)
FEATURES             Location/Qualifiers
     source          1..1894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             415..1423
                     /note="human JKETS."
                     /codon_start=1
                     /protein_id="CAC19002.1"
                     /db_xref="GI:11876968"
                     /translation="MISASPLGSVSESHLLEPGVSPSLSPEAAACAVSLEEFDWS
                     FSFPATPEQGLSAFYLSYFDMGVPFDSSWAAKADCASSREFPPFFEQCLVTDSQAPA
                     GSLDLVPGGLPLEEHSLEQVQSMVVGEVLKDTETACKLLNLTADEMDWSPSNVQKWLL
                     WIEHQYKLPFMGKAFQELAGKELCAMSEEQFKQKSPLGGFVLHAHLTQWESAAWMKEE
                     TSPGATHYCASISEESWIDSEVDSSCSGQVTHLWGFLKELLLKPHSYGRFIRWLNKEK
                     GILKIPLSALVARLWGIEENFPAMNYDKLSRSIPQYYKNGIIEKPDISQKLVYQFVHP
                     I"
BASE COUNT       368 a    653 c    571 g    302 t
ORIGIN


alignment scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_38_46 x AX048148  ..

 Align seq 1/2  to: AX048148  from: 1  to: 1894

       1 LeuGluArgArgAspIleSerProSer 9
         ||||||||||||||||||||||||||
     527 CTCGAGAGACGGGACTCGGAGTCCCAGT 553

seq_name: qb_pr4:AF071538

seq_documentation_block:
LOCUS       AF071538      1894 bp    mRNA            PRI       16-JAN 2000
DEFINITION  Homo sapiens Ets transcription factor PDEF (PDEF) mRNA, complete
            cds.
ACCESSION   AF071538
VERSION     AF071538.1  GI:4007417
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1894)
  AUTHORS   Oettgen,P., Finger,E., Sun,Z., Akbarali,Y., Thamrongsak,U.,
```

```
            Boltax,I., Grall,F., Dabo,A., Weiss,A., Brown,L., Quinn,G., Kas,K.,
            Endress,G., Kunsch,C. and Libermann,T.A.
  TITLE     PDEF, a novel prostate epithelium specific ets transcription
            factor, interacts with the androgen receptor and activates
            prostate specific antigen gene expression
  JOURNAL   J. Biol. Chem. 275 (2), 1216-1225 (2000)
  MEDLINE   20092894
REFERENCE   2  (bases 1 to 1894)
  AUTHORS   Libermann,T.A., Oettgen,P., Kunsch,C. and Finger,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (10 JUN 1998) Medicine, Beth Israel Deaconess Medical
            Center, 330 Brookline Ave., Boston, MA 02215, USA
FEATURES             Location/Qualifiers
     source          1..1894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6p21.3"
                     /tissue_type="prostate"
     gene            1..1894
                     /gene="PDEF"
     CDS             416..1423
                     /gene="PDEF"
                     /note="prostate epithelium specific"
                     /codon_start=1
                     /product="Ets transcription factor PDEF"
                     /protein_id="AAC95299.1"
                     /db_xref="GI:4754126"
                     /translation="MGSASHCLSSVSFSNLLLGEPDLVSMTGLEKAAACAVCLERRDWS
                     PSPPAIPEQGLSAFYLSYFDMLYPEDSSWAAKAPGASSKEEPEEPEEPEQCPVTDSQAPA
                     GSLDLVPGGLTLEEHSLEQVQSMVVGEVLKDTEIACKLLNLTADPMDWSPSNVQKWLL
                     WIEHQYRLPPMGKAEQELAGKELQAMSEEEQFPQKSPLGGPVLRAHLDTRWKSAAWMEER
                     TSPCAIHYCASISEESWTDSEVDSSCSGQPTHLWQFLKNELLELEPHSYCPFTEWLNKEK
                     GIFKIEDSAQVARLWGTRKNRPTAMNYDKLSRSLRYYYERGIMRKPDISQRLVYQFVHP
                     I"
BASE COUNT       368 a    654 c    571 g    302 t
ORIGIN


alignment scores:
            Quality:     9.00              Length:       9
              Ratio:    1.000               Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-9458-2_COPY_38_46 x AF071538  ..

 Align seq 1/1  to: AF071538  from: 1  to: 1894

        1 LeuGluArgArgAspTrpSerProSer 9
          |||||||||||||||||||||||||||
      527 CTCGAGAGACGGGACTGGAGTCCCAGT 553

seq_name: qb_pri:AB031549

seq_documentation_block:
LOCUS       AB031549      1898 bp    mRNA            PRI       28-JAN-2000
DEFINITION  Homo sapiens PSE mRNA for prostate ets, complete cds.
ACCESSION   AB031549
VERSION     AB031549.1  GI:6721497
KEYWORDS    prostate ets.
SOURCE      Homo sapiens adult male prostate epithelium prostate cancer
            cell line:PC-3 cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Yamada,N., Tamai,Y., Miyamoto,H. and Nozaki,M.
  TITLE     Cloning and expression of the mouse Pse gene encoding a novel Ets
            family member
  JOURNAL   Gene 241 (2), 267-274 (2000)
  MEDLINE   20147504
REFERENCE   2  (bases 1 to 1898)
  AUTHORS   Nozaki,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2000) To the DDBJ/EMBL/GenBank databases. Masami
            Nozaki, Osaka University, Institute for Microbial Diseases; 3-1,
            Yamadaoka, Suita, Osaka 565-0871, Japan
            (E-mail:pse-ets@biken.osaka-u.ac.jp, Tel:81-6-879-8338,
            Fax:81-6-879-8339)
FEATURES             Location/Qualifiers
     source          1..1898
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="PC-3"
                     /cell_type="prostate cancer"
                     /chromosome="6"
                     /dev_stage="adult"
                     /map="6p21.2-21.3"
                     /sex="male"
                     /tissue_type="prostate epithelium"
                     /vector="vector:pBC18"
     gene            406..1413
                     /gene="PSE"
     CDS             406..1413
                     /gene="PSE"
                     /codon_start=1
                     /product="prostate ets"
                     /protein_id="BAA89543.1"
                     /db_xref="GI:6721498"
                     /translation="MGSASHCLSSVSFSNLLLGEPDLVSMTGLEKAAACAVCLERRDWS
                     PSPPAIPEQGLSAFYLSYFDMLYPEDSSWAAKAPGASSKEEPEEPEEPEQCPVTDSQAPA
                     GSLDLVPGGLTLEEHSLEQVQSMVVGEVLKDTEIACKLLNLTADPMDWSPSNVQKWLL
                     WIEHQYRLPPMGKAEQELAGKELQAMSEEEQFPQKSPLGGPVLRAHLDTRWKSAAWMEER
                     TSPCAIHYCASISEESWTDSEVDSSCSGQPTHLWQFLKNELLELEPHSYCPFTEWLNKEK
                     GIFKIEDSAQVARLWGTRKNRPTAMNYDKLSRSLRYYYERGIMRKPDISQRLVYQFVHP
                     I"
BASE COUNT       380 a    651 c    567 g    300 t
ORIGIN


alignment scores:
            Quality:     9.00              Length:       9
              Ratio:    1.000               Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-9458-2_COPY_38_46 x AB031549  ..

 Align seq 1/1  to: AB031549  from: 1  to: 1898

        1 LeuGluArgArgAspTrpSerProSer 9
          |||||||||||||||||||||||||||
      517 CTCGAGAGACGGGACTGGAGTCCCAGT 543

seq_name: qb_pat1:AX048159

seq_documentation_block:
LOCUS       AX048159      3317 bp    DNA                PAT       15-DEC-2000
DEFINITION  Sequence 12 from Patent WO0070092.
ACCESSION   AX048159
VERSION     AX048159.1  GI:11876975
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3317)
  AUTHORS   Kaufmann,J. and Xin,H.
  TITLE     Expression of ets domain proteins in cancer
  JOURNAL   Patent: WO 0070092-A 12-23-NOV-2000;
            Chiron Corporation (US)
FEATURES             Location/Qualifiers
     source          1..3317
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

```
                     /note="GSEF encoding sequence with promoter (Plac -2A-2B)"
     promoter        1. .1380
                     /n. te="Pr.m. ..."
     CDS             1815. .2822
                     /note="GSEF coding sequence"
                     /codon_start=1
                     /protein_id="CAC19004.1"
                     /db_xref="GI:11876976"
                     /translation="MESASIGLSSVLGSMLLGFLVSRLGLKAAVGAVGLEKRAMS
                     PSPPAIPEQGLSAFYLSYFIMLYFFDSSWAAKAIGASSFLEIFEETAFIVIISQAIA
                     GSLDLVPGGLTLEEHSLEQVQSMVVGFVLKDIPTACFLINLSAFMPAVFSNVLPWLI
                     WIEHQYPLPPMCKAPQELACFFLCAMSFPLFRQKFPLGTVFRAPLTWFCAAMMPLS
                     TSPGAIHYCASTSPFSWTDSFVDSSCSQQFDHIWQFLKFLLELPHSYGPFIPWLNKFK
                     GIFFFILSA.VASIWSIPFNFLAMNFLFLFLFLVTDEFTVLPVSI
                     1"
BASE COUNT      710 a   1026 c    970 g    611 t
ORIGIN


alignment_scores:
            Quality:     9.00              Length:       9
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_38_46 x AX048159   ..

 Align seq 1/1  to: AX048159  from: 1  to: 3317

        1 LeuGluArgArgAspTrpSerProSer 9
          |||||||||||||||||||||||||||
     1926 CTCGAGACACGGCACTGGAGTCCCAGT 1952

seq_name: qb_pr6:AL157372

seq_documentation_block:
LOCUS        AL157372   150660 bp    DNA             PRI       29-NOV-2000
DEFINITION   Human DNA sequence from clone RP11-375E1 on chromosome 6, complete
             sequence.
ACCESSION    AL157372
VERSION      AL157372.18  GI:11493236
KEYWORDS     HTG.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 150660)
  AUTHORS    Skuce,C.
  TITLE      Direct Submission
  JOURNAL    Submitted (29-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
             CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
             requests: clonerequest@sanger.ac.uk
COMMENT      On Nov 30, 2000 this sequence version replaced gi:11876976.
             During sequence assembly data is compared from overlapping clones.
             Where differences are found these are annotated as variations
             together with a note of the overlapping clone name. Note that the
             variation annotation may not be found in the sequence submission
             corresponding to the overlapping clone, as we submit sequences with
             only a small overlap as described above.
             This sequence has been finished according to sequence map criteria
             as follows. An attempt is made to resolve all sequencing problems,
             such as compressions and repeats, but not necessarily within known
             annotated repeat sequence elements. Where the sequence is
             ambiguous, there is an annotation using the 'unsure' feature key.
             The following abbreviations are used to associate primary accession
             numbers given in the feature table with their source databases:
             EM:, EMBL; SW:, SWISSPROT; IN:, TREMBL; WP:, WORMPEP; Information
             on the WORMPEP database can be found at
             http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
             was generated from part of bacterial clone contigs of human
             chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
             Group. Further information can be found at
             http://www.sanger.ac.uk/HGP/Chr6
             RP11-375E1 is from the library RPCI-11.2 constructed at the Roswell
             Park Cancer Institute by the group of Pieter de Jong. For further
             b.t.'' .. http://bacpac.med.buffalo.edu/
             VECTOR: pBACe3.6
             IMPORTANT: this sequence is not the entire insert of clone
             RP11-375E1 it may be shorter because we sequence overlapping
             sections only once, except for a 100 base overlap.
             The true left end of clone RP11-375E1 is at 1 in this sequence. The
             true left end of clone RP3-391O22 is at 150561 in this sequence.
             The true right end of clone RP11-302L10 is at 11044 in this
             sequence.
FEATURES     Location/Qualifiers
     source  1. .150660
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="6"
             /clone="RP11-375E1"
             /clone_lib="RPCI-11.2"
BASE COUNT    45470 a  38867 c  38740 g  47583 t
ORIGIN


alignment_scores:
            Quality:     9.00              Length:       9
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_38_46 x AL157372/rev   ..

 Align seq 1/1  to reverse of: AL157372  from: 1  to: 150660

        1 LeuGluArgArgAspTrpSerProSer 9
          |||||||||||||||||||||||||||
   129629 CTCGAGACACGGCACTGGAGTCCCAGT 129603

seq_name: qb_pri:AC005192

seq_documentation_block:
LOCUS        AC005192   128600 bp    DNA             PRI       03-FEB-2000
DEFINITION   Homo sapiens BAC clone CIB 163K11 from 7q31, complete sequence.
ACCESSION    AC005192
VERSION      AC005192.1  GI:4264575
KEYWORDS     HTG.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 128600)
  AUTHORS    Kalicki,J. and Gibson,A.
  TITLE      The sequence of Homo sapiens BAC clone CIB-163K11
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 128600)
  AUTHORS    Waterston,R.
  TITLE      Direct Submission
  JOURNAL    Submitted (27-JUN-1998) Department of Genetics, Washington
             University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE    3  (bases 1 to 128600)
  AUTHORS    Waterston,R.
  TITLE      Direct Submission
  JOURNAL    Submitted (03-FEB-2000) Department of Genetics, Washington
             University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT      ---------- Genome Center
               Center: Washington University Genome Sequencing Center
               Center code: WUGSC
               Web site: http://genome.wustl.edu
               Contact: sapiens@watson.wustl.edu
             ---------- Summary Statistics
               Center project name: H_RG163K11
             ---------- ---------

             NOTICE: This sequence may not represent the entire insert of this
             clone. It may be shorter because we only sequence overlapping
```

```
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone; and the assembly was
            confirmed by restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center.  For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/..., send
            mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/...

            SOURCE INFORMATION:
            Clone CTB-163K11 is from the first release of the human BAC library
            CITB-978SK-B.  The library contains cloned DNA from the male
            fibroblast cell line 978SK.  See:  Shizuya et al., Proc. Natl.
            Acad. Sci. USA 89:8794-7 (1992); H.J. Kim et al., Genomics 34:213-8
            (1996).  This clone is available from Research Genetics, Inc.
            (http://www.resgen.com).
            VECTOR:  pBeloBAC11
            Selection:  chloramphenicol

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the right is CTA-302F4, 200 bp overlap.
            Actual start of this clone is at base position 1 of CTB-163K11;
            actual end is at 29763 of CTA-302F4.
FEATURES             Location/Qualifiers
     source          1..128600
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q31"
                     /clone="CTB-163K11"
                     /clone_lib="CITB-978SK-B"
     misc_feature    390..527
                     /note="match to EST W88459 (NID:gI404811) zh72e04.r1"
     misc_feature    468..527
                     /note "match to EST W87947 (NID:gI402031) zh68109.s1"
     gene            476..14129
                     /gene "WUGSC:H_RG163K11.1"
     CDS             join(476..526,611..789,880..988,6591..6725,10829..10957,
                     11476..11471,14040..14129)
                     /gene "WUGSC:H_RG163K11.1"
                     /note "similar to mouse interferon related protein IFI4;
                     96% identical to PI9182 (PID:gI35861); H_RG163K11.1"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAC24562.1"
                     /db_xref="GI:3264576"
                     /translation="CAIGFGVCCFIAIIGCIIELYSILECLENIFIKSSLKEKDLIVIC
                     VFPNFVFHFSSFLAWTLIFPCPLNFVFFFLFMHIFRFCFLSCHFVNMSIAACEGCA
                     LLFFELARGIESPFFYEDMFSLFQMLPALAIICNFHPAFVFFFFQRSVFRDVLRAVEER
                     DFPTFLIFFCPFEEMYIFCWVFFHIYFIFKIVIGCGMGLGIGSNFLLFNVFLLGFFVML
                     DAATLKIMKISPFFPHLYNSAAFFAGIFAPSFRDFRADVGFF"
     misc_feature    476..527
                     /gene "WUGSC:H_RG163K11.1"
                     /note "match to EST AA968021 (NID:gI4141914) nh07g02.r1"
     misc_feature    478..527
                     /gene "WUGSC:H_RG163K11.1"
                     /note "match to EST W88459 (NID:gI404811) zh72e04.s1"
     misc_feature    602..790
                     /gene "WUGSC:H_RG163K11.1"
                     /note "match to EST W87947 (NID:gI402031) zh68109.s1"
     misc_feature    602..790
                     /gene "WUGSC:H_RG163K11.1"
                     /note "match to EST W88459 (NID:gI404811) zh72e04.r1"
     misc_feature    602..790
                     /gene "WUGSC:H_RG163K11.1"
                     /note "match to EST W87946 (NID:gI402030) zh68109.r1"
     misc_feature    620..787
                     /gene "WUGSC:H_RG163K11.1"
                     /note "match to EST AA968021 (NID:gI4141914) nh07g02.r1"
     misc_feature    complement(690..790)
                     /note "match to EST W87946 (NID:gI402030) zh68109.r1"
     misc_feature    740..790
                     /gene "WUGSC:H_RG163K11.1"
                     /note "match to EST AA298817 (NID:gI951170)"
     misc_feature    880..989
                     /gene "WUGSC:H_RG163K11.1"
                     /note "match to EST W88459 (NID:gI404811) zh72e04.r1"
     misc_feature    complement(880..989)
                     /note "match to EST W87946 (NID:gI402030) zh68109.r1"
     misc_feature    880..989
                     /gene "WUGSC:H_RG163K11.1"
                     /note "match to EST AA298817 (NID:gI951170)"
     misc_feature    880..933
                     /gene "WUGSC:H_RG163K11.1"
                     /note "match to EST AA968021 (NID:gI4141914) nh07g02.r1"
     misc_feature    880..989
                     /gene "WUGSC:H_RG163K11.1"
                     /note "match to EST W87947 (NID:gI402031) zh68109.s1"
     misc_feature    888..989
                     /gene "WUGSC:H_RG163K11.1"
                     /note "match to EST H03717 (NID:gI866650) yj43a02.r1"
     repeat_region   1542..1654
                     /rpt_family "MER81"
     repeat_region   2400..2504
                     /rpt_family "(CATA)n"
     repeat_region   3042..4181
                     /rpt_family "MIR"
     repeat_region   3656..4822
                     /rpt_family "Alu"
     repeat_region   5018..5238
                     /rpt_family="MER1_type"
     misc_feature    6591..6724
                     /gene "WUGSC:H_RG163K11.1"
                     /note="match to EST W87947 (NID:gI402031) zh68109.s1"
     misc_feature    6591..6726
                     /gene "WUGSC:H_RG163K11.1"
                     /note "match to EST AA298817 (NID:gI951170)"
     misc_feature    complement(6591..6726)
                     /note "match to EST W87946 (NID:gI402030) zh68109.r1"
     misc_feature    complement(6591..6726)
                     /note "match to EST AA678021 (NID:gI2631520) aa59b12.s1"
     misc_feature    6591..6726
                     /gene "WUGSC:H_RG163K11.1"
                     /note "match to EST H03717 (NID:gI866650) yj43a02.r1"
     misc_feature    6660..6726
                     /gene "WUGSC:H_RG163K11.1"
                     /note "match to EST N41796 (NID:gI1076827) yx98a...r1"
     repeat_region   7118..7402
                     /rpt_family "Alu"
     repeat_region   8971..9168
                     /rpt_family "L2"
     repeat_region   9251..9562
                     /rpt_family "MER2_type"
     repeat_region   9763..9938
                     /rpt_family "MIR"
     misc_feature    10819..10866
                     /gene "WUGSC:H_RG163K11.1"
                     /note "match to EST W87947 (NID:gI402031) zh68109.s1"
     misc_feature    complement(10827..10957)
                     /note "match to EST AA678021 (NID:gI2631520) aa59b12.s1"
     misc_feature    10827..10879
                     /gene "WUGSC:H_RG163K11.1"
                     /note "match to EST AA298817 (NID:gI951170)"
     misc_feature    10827..10957
                     /gene "WUGSC:H_RG163K11.1"
```

```
                     /note "match to EST H03717 (NID:g866650) yj43a02.r1"
     misc_feature    10827. .10957
                     /note "WashU-R pc[illegible]"
                     /note "match to EST N41796 (NID:g1165827) yw88a03.r1"
     misc_feature    complement(10827. .10949)
                     /note "match to EST W87946 (NID:g1402830) zh68f09.r1"
     misc_feature    complement(10829. .10957)
                     /note "match to EST AA705511 (NID:g2711470) [illegible]"
     misc_feature    complement(10847. .10957)
                     /note "match to EST H29848 (NID:g869548) yn79e05.s1"
     misc_feature    complement(10875. .10957)
                     /note "match to EST AA941101 (NID:g2696389) [illegible]"
     misc_feature    complement(10895. .10957)
                     /note "match to EST AA994664 (NID:g2694602) zl50002.s1"
     misc_feature    complement(11374. .11471)
                     /note "match to EST AA628021 (NID:g2643543) [illegible]"
     misc_feature    11374. .11471
                     /gene "WUGSC:H_RG164K11.1"
                     /note "match to EST N41796 (NID:g1165827) yw88a03.r1"
     misc_feature    complement(11374. .11471)

alignment_scores:
            Quality:     8.00              Length:       8
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_38_46 x AC005192/rev   .

 Align seq 1/1  to reverse of: AC005192  from: 1  to: 128600

       1 LeuGluArgArgAspTrpSerPro 8
         ||||||||||||||||||||||||
   46048 TTAGAAAGAAGAGACTGGAGTCCC 46015

seq_name: qb_pr7:BC002903

seq_documentation_block:
LOCUS       BC002903     1743 bp    mRNA            PRI       16-MAR-2001
DEFINITION  Homo sapiens, clone MGC:10323, mRNA, complete cds.
ACCESSION   BC002903
VERSION     BC002903.1  GI:12804100
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1743)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-FEB-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: CHTN/NIH
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium [LLNL]
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site:       http://www.nisc.nih.gov/
            Contact:        nisc_mgc@mail.nih.gov
            Shevchenko,Y., Wetherby,K.D., Beckstrom-Sternberg,S.M.,
            Benjamin,R., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Guan,X., Gupta,J., Ho,S.-L., Karlins,E., Lopaspi,E.,
            Lim,M., Maduro,Q.L., Masiello,C., Mastrian,S.D., McCloskey,J.C.,
            McDowell,J., Pearson,E., Snyder,B., Stantripop,S., Thomas,P.J.,
            Tiongson,E.E., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
            Zhang,L. H. and Green,E.D.
            Clone distribution: MGC clone distribution information can be found
            [illegible] /image=6141396" FEATURES
            Series: IRAL Plate: 15 Row: m Column: 21.
FEATURES             Location/Qualifiers
     source          1. .1743
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="MGC:10323 IMAGE:6944615"
                     /tissue_type="Lung, small cell carcinoma"
                     /clone_lib="NIH_MGC_7"
                     /lab_host="DH10B:R"
                     /note "Vector: pOTB7"
     CDS             816. .956
                     /codon_start=1
                     /product="Unknown (protein for MGC:10323)"
                     /protein_id="AAH02903.1"
                     /db_xref="GI:12804101"
                     /translation="MASGKLIGVEHISKYGLDAKFIKMLVELMQSHAVHHLFIVH
                     LR"
BASE COUNT      423 a    537 c    455 g    328 t
ORIGIN


alignment_scores:
            Quality:     7.00              Length:       7
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_38_46 x BC002903/rev   ..

 Align seq 1/1  to reverse of: BC002903  from: 1  to: 1743

       8 ArgArgAspTrpSerProSer 9
         ||||||||||||||||||||
    1391 CGCAGGGACTGGTCCTCCTCT 1371

seq_name: qb_rn2:RATENKA4

seq_documentation_block:
LOCUS       RATENKA4     2122 bp    DNA             ROD       27-APR-1993
DEFINITION  Rat dynorphin gene, exon 4.
ACCESSION   M32784
VERSION     M32784.1  GI:204036
KEYWORDS    alpha-neoendorphin; dynorphin.
SEGMENT     4 of 4
SOURCE      Rat DNA, and cDNA to mRNA, clone PLRD3.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 2122)
  AUTHORS   Douglass,J., McMurray,C.T., Garrett,J.E., Adelman,J.P. and
            Calavetta,L.
  TITLE     Characterization of the rat prodynorphin gene
  JOURNAL   Mol. Endocrinol. 3, 2070-2078 (1989)
  MEDLINE   90190668
FEATURES             Location/Qualifiers
     source          1. .2122
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
     CDS             join(M32783.1:66, .197,51, .665)
                     /note "prodynorphin precursor"
                     /codon_start=1
                     /protein_id="AAA41117.1"
                     /db_xref="GI:204038"
                     /translation="MAWSRLMLAACLLVLPSEVAALQLSLSEAVEKESGEWFFNFL
                     DQSFPQQELVPPSEEWETCRGFWSPLTLTASCLPCKLDENEVALEDCYLAFKKLFP
                     LLKELLEKGQLLLSVSLEKLRGLSSRLGNGRLSELLGTLEMNDEAALALLHFNEEDLK
                     KLAEKYGGFLRKQFKVVTRSQENPNTYSEDLDV"
     intron          1. .50
```

```
                     /note "dynorphin intron C"
     prim_transcript  1..2122
                     /note "dynorphin mRNA and introns"
     exon             51..665
                     /note "prodynorphin precursor"
                     /number 4
     mat_peptide      414..440
                     /note "alpha-neoendorphin"
     mat_peptide      522..572
                     /note "dynorphin A"
     mat_peptide      579..662
                     /note "dynorphin B"
BASE COUNT       546 a    528 c    544 g    514 t
ORIGIN      About 2 Kb after segment 3.


alignment scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2_COPY_38_46 x RATENKA4

 Align seq 1/1  to: RATENKA4  from: 1  to: 2122

        3 ArgArgAspTrpSerProSer 9
          ||||||||||||||||||||||
     1667 AGGAGAGACTGGTCCCCCTCG 1687

seq name: gb pr5:AK026996

seq_documentation_block:
LOCUS        AK026996      2784 bp    mRNA            PRI       29-SEP-2000
DEFINITION   Homo sapiens cDNA: FLJ23444 fis, clone HEP14562.
ACCESSION    AK026996
VERSION      AK026996.1  GI:10439996
KEYWORDS     oligo capping; fis (full insert sequence).
SOURCE       Homo sapiens hepatoma cell_line:HepG2 cDNA to mRNA, clone lib:HEP
             clone:HEP14562.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (sites)
  AUTHORS    Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
             Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
             Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
  TITLE      NEDO human cDNA sequencing project
  JOURNAL    Unpublished (2000)
REFERENCE    2  (bases 1 to 2784)
  AUTHORS    Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
             Shibahara,T., Tanaka,T. and Nakamura,Y.
  TITLE      Direct Submission
  JOURNAL    Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio
             Sugano, Institute of Medical Science, University of Tokyo,
             Laboratory of Genome Structure Analysis, Human Genome Center;
             Shirokane dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
             (E-mail:ednal@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
             Fax:81-3-5449-5416)
COMMENT      NEDO human cDNA sequencing project supported by Ministry of
             International Trade and Industry of Japan; cDNA full insert
             sequencing: Research Association for Biotechnology; cDNA library
             construction, 5'- & 3'-end one pass sequencing: Department of
             Virology and Human Genome Center, Institute of Medical Sciences,
             University of Tokyo (partly supported by Science and Technology
             Agency).
FEATURES             Location/Qualifiers
     source          1..2784
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="HepG2"
                     /cell_type="hepatoma"
                     /clone="HEP14562"
                     /clone_lib="HEP"
                     /note="cloning vector pME18SFL3"
BASE COUNT       723 a    633 c    603 g    825 t
ORIGIN


alignment scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2_COPY_38_46 x AK026996  ..

 Align seq 1/1  to: AK026996  from: 1  to: 2784

        1 LeuGluArgArgAspTrpSer 7
          |||||||||||||||||||||
     1081 CTGGAAAGGAGAGACTGGAGC 1101

seq name: gb pr2:RATENKB

seq_documentation_block:
LOCUS        RATENKB       2963 bp    DNA             ROD       27-APR-1993
DEFINITION   Rat prodynorphin (Preproenkephalin B) gene, major (3') exon.
ACCESSION    M10088
VERSION      M10088.1  GI:204039
KEYWORDS     Leu enkephalin; beta neoendorphin; leumorphin; preprodynorphin;
             prodynorphin.
SOURCE       Rat liver DNA, clone p-lambda-RD2 19.
  ORGANISM   Rattus norvegicus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae;
             Rattus.
REFERENCE    1  (bases 1 to 2963)
  AUTHORS    Civelli,O., Douglas,J., Goldstein,A. and Herbert,E.
  TITLE      Sequence and expression of the rat prodynorphin gene
  JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 82, 4291-4295 (1985)
  MEDLINE    85216080
COMMENT      A draft entry and copy of this sequence were kindly supplied
             diskette by O Civelli (21 OCT-1985).
             Polyadenylation signals are present at the following positions:
             2061-2066, 2441-2446, 2445-2450, 2780-2785, 2867-2872.  The mRNA 3'
             end was assigned by analogy to the known polyadenylation sites in
             human and porcine prodynorphin genes [1].
FEATURES             Location/Qualifiers
     source          1..2963
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
     prim_transcript 1..2462
                     /note="prodyn mRNA"
     gene            1..890
                     /gene="dynorphin"
     intron          1..390
                     /gene="dynorphin"
                     /number=1
     CDS             391..1005
                     /partial
                     /codon_start=1
                     /product="dynorphin"
                     /protein_id="AAA41118.1"
                     /db_xref="GI:204040"
                     /translation="MCSLECQDLVPPSEEWEICKGFWSEFTLTASGLDGKMELENEVA
                     LLEGYIALIKLLEPLLKEIENGQLLTSVSELKRGLSSSEGNGKLSELLGTLEMNDEA
                     AQAVTLDLNELELWEQAKDYGGFLRKEYEKPKEDLMACTEEDPQQTQCPAVCHEFLYERYGG
                     FLERIRPKLKWDNQKKYGGFLRRQFKVVTKSQENPNTYSEDLDV"
     mat_peptide     754..768
                     /product="Leu enkephalin"
     mat_peptide     754..780
                     /product="beta-neoendorphin"
     mat_peptide     862..877
                     /product="Leu enkephalin"
     mat_peptide     862..903
```

```
                     /product "dynorphin 1-17"
     mat_peptide     919. .932
                     /product "Leu-enkephalin"
     mat_peptide     959. .1002
                     /product "leumorphin"
BASE COUNT       765 a    717 c    748 g    743 t
ORIGIN      1 bp upstream of EcoRI site.


alignment_scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_38_46 x RATENKB    .

 Align seq 1/1  to: RATENKB  from: 1  to: 2069

        8 ArgArgAspTrpSerProSer 9
          ||||||||||||||||||||||
     2007 AGGAGAGACTGGTCCCCTCG 2027

seq_name: gb_pri:AB046773

seq_documentation_block:
LOCUS       AB046773      5538 bp    mRNA             INI       22-FEB-2001
DEFINITION  Homo sapiens mRNA for KIAA1553 protein, partial cds.
ACCESSION   AB046773
VERSION     AB046773.1  GI:10047170
KEYWORDS    .
SOURCE      Homo sapiens cDNA to mRNA, clone:thl7011.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Nagase,T., Kikuno,R., Nakayama,M., Hirosawa,M. and Ohara,O.
  TITLE     Prediction of the coding sequences of unidentified human genes.
            XVIII. The complete sequences of 100 new cDNA clones from brain
            which code for large proteins in vitro
  JOURNAL   DNA Res. 7 (4), 273-281 (2000)
   PUBMED   10997877
REFERENCE   2  (bases 1 to 5538)
  AUTHORS   Ohara,O., Nagase,T. and Kikuno,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-2000) Osamu Ohara, Kazusa DNA Research Institute,
            Department of Human Gene Research, 1532-3, Yana, Kisarazu, Chiba
            292-0812, Japan (E-mail:ednfinfo@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/hgene, Tel:81-438-52-3933,
            Fax:81-438-52-3914)
FEATURES             Location/Qualifiers
     source          1. .5538
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="thl7011"
                     /note "vector:pBluescript II SK plus"
     gene            1. .2014
                     /gene="KIAA1553"
     CDS             <1. .2014
                     /gene="KIAA1553"
                     /note "Start codon is not identified."
                     /codon_start=2
                     /product="KIAA1553 protein"
                     /protein_id="BAB13879.1"
                     /db_xref="GI:10047171"
                     /translation="ANASN[illegible]
                     LNTLTSNMSQLHSKVQLL[illegible]
                     LSGGDIA[illegible]
                     KDTAVWQAECLPQLNDFFSKFWAQREMPDSQPSCQVASFFFAPLQVDPGHELTNKFQFE
                     ALSLDRSSTIASDHVVDIQGLIEELDEASSPGDEAVFLHELFFELFDHEKLQEQYSC
                     YGDGGKQFLQPQRLQLIRNYTFIYFPDMQFFAWIQQCAQRLNDPLPGHSLDAASFPVD
                     PPFDDCYLSSAGFPSCSVVFVIGSFSGEHEGGRSKEIWLVFLPELELCPLPDFEVIC
                     ADCLLSKFFQLRSLYPSSLSLQNPASPLIVHLPPELFTHPNLPFQYNCSCSLCKEQLDP
                     SPIRLIHHYVQLLYPRAKNQPVWILEFVGKLDSPCPRFFLQQRESYQLLSKVHVPGLE
                     PRQLISYAINEFRFKEEFEQPLPPERSSAIPCNLIFLQLVVPSCGEVFSPYLLSQK
                     LVEELVQQSLSVGNLAAELGVRLLPLLLTALNLRLQYNIGLACNEEQLLPTRLRLIRH
                     YVPAVSPVFPMFLVWEYCLPSIDFPPPPNPREPCTQLLEEAEKVEE"
BASE COUNT       1305 a    1486 c    1382 g    1365 t
ORIGIN


alignment_scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_38_46 x AB046773    .

 Align seq 1/1  to: AB046773  from: 1  to: 5538

        1 LeuGlnArgArgAspTrpSer 7
          |||||||||||||||||||||
     3844 CTGGAAAGGAGGGACTGGAGC 3864

seq_name: gb_ba3:RFCGCP

seq_documentation_block:
LOCUS       RFCGCP         7304 bp    DNA             BCT       27-APR-1994
DEFINITION  Ruminococcus flavefaciens DNA sequence of cellulase gene complex to:
            endo-glucanase, exo-glucanase, beta-glucosidase and protease.
ACCESSION   X56082
VERSION     X56082.1  GI:46152
KEYWORDS    beta-glucosidase; glucanase; protease.
SOURCE      Ruminococcus flavefaciens.
  ORGANISM  Ruminococcus flavefaciens
            Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
            Ruminococcus.
REFERENCE   1  (bases 1 to 7304)
  AUTHORS   Huang,C.M., Asmundson,R.V. and Yu,P.L.
  TITLE     Nucleotide sequence of a cellulase gene complex from Ruminococcus
            flavefaciens strain 186 coding for multi-cellulase activities
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 7304)
  AUTHORS   Huang,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-SEP-1990) C.M. Huang, DSIR FRUIT AND TREES, PRIVATE
            BAG, PALMERSTON NORTH, NEW ZEALAND
FEATURES             Location/Qualifiers
     source          1. .7304
                     /organism="Ruminococcus flavefaciens"
                     /strain "186"
                     /db_xref "taxon:1265"
                     /clone_lib "lambda CM series constructed with EcoRI site
                     in lambda NM1149 vector"
                     /clone "lambda CM903"
     CDS             829. .2871
                     /codon_start=1
                     /transl_table=11
                     /product "endo-glucanase"
                     /protein_id "CAA39559.1"
                     /db_xref "GI:46153"
                     /db_xref "SPTREMBL:Q52747"
                     /translation="MVTSFSARSGTPSSSPTSPSLSFSGTTFTSPSFSSSPTSSTA
                     SFSSPTVSCKEAVAFACHADARTAGKKDLKLQCDAPLPCKHHQFKLSHAVVKEDGTGI
                     RQKDCVRGACHGPSVHRADWHGFKRAALIEVGGLVHDQGKLVPVVRGVPAPPPRDKRR
                     DQSKFRKALVQLPLPRDREQPGGHQPHDADAVHGGGRGVRGQLQSKKLKKDKFVRDHRD
                     GPSDCRSFDPDPDRQRAVRKQRGRRHRLVDKRKVRPREQPVPKVRQKQCPVKAHEHKKKR
                     GLAGGLVSEAQAAHPKSVQKELDHGGKECHKVCAGGAGAKCACDCDAVKEEKGKAQR
                     IKKCGRLKKKAKCGKEASIGLKRAGVPSGLQKRQASAAVLQHHPGKKFLPIRGLQKEAV
                     KSLKAIQAGASQLKKVCPKGGVCKLKVSALKVVLLKKKSCKDKALCDAKLCVSHRAI
                     PPCHPHPCPAVRCSPQSDPARCLRRRHVDAIAQGQPRDPPQLPVLVHLIERKAIGQRI
                     PAMGARAHPITHWSGVKLPRGARASQGHRRDRCIAVKLPPVRRRSKAQIRIKAIVVRARR
                     IPTPSCACTPFAFLPPVHPPPPRPAIVGPVIVPCPFFPVKSDFHKSHGTLKAQHVHP
                     ALKDAAGVRAGIGLQKVHKKAAAIKGIGKQGIKKGSGGVGGAGKGGATASGQIIA"
     misc_feature    2072
```

```
                     /note "start codon according to author"
     CDS             2072. .3376
                     /note "Protein sequence is in conflict with the conceptual
                     translation"
                     /codon_start 1
                     /transl_table 11
                     /product "beta-glucosidase"
                     /protein_id "CAA89560.1"
                     /db_xref "GI:1333803"
                     /db_xref "SPTREMBL:Q52748"
                     /translation "SAGYASLRYVSSPSGAGLDAIVRIVIIGKGSPIGLLKAGTRIGV
                     GLCGAPGKMFVRETYGAVTWTRRRAKVGGIGPSYGYWYTNCGGKLRASVKTLWELART
                     RSTGAVSGYDAGRGLAKIFGGTALLSGYPASAATVVPNFTLAPTSSAHVGYVRAGALP
                     VLSELSYVASTSGAAVALRWAALRSEAASGSASEATPTAATVKAPNMSTELLATLPGC
                     ERASANNASIGAPKRPEAPADKKKADRARATRPARDAGEPSLAAKRQPDLYRTNSGYK
                     ALARVRSLRVQRYASVASFRTVACASRAGSTERDRATRAIVSTFTDPSSTIVCSIVRVV
                     RTVRAYLPLSTRQAASYSKAIASSDRVELYYTSRYTYGGRRAPTTALPYSANLRSVYY
                     ASYALSVEAIQRGLRYVYARTRVIWRVLDDFIYSYARTGGSA"
     CDS             4586. .5304
                     /codon_start 1
                     /transl_table 11
                     /product "exo-glucanase"
                     /protein_id "CAA89561.1"
                     /db_xref "GI:46155"
                     /db_xref "SPTREMBL:Q52749"
                     /translation "MRQDTQSAQREPVHDLDLDPVSGVLRLASDEQHGRSVDCLDQHV
                     AVEGRDHDLSAHDLLGVAARVSDEPLLYRDDPADREQPRPDGRQAVPFPVDGCAASDC
                     HDLRVDALQHSGLSQQHDRGRPEVARHLDEDGKIVLRADSQDDTQDSRLQLLAWCRLL
                     TVRGHVDPVVAHTGRQADEQDAGDELPSVHQAEDGGALSRQPDHLDLNVSRDDHSVER
                     RLDQDSHDLQRALQDQDDRLAAVLRLDGDHLDDPVGVQLPRQDHDDSAQDHRSQGERLA
                     SGHHPDLRSFQVTAALLEDEAHVVGHQASSTRVDSPDQERGDHPDRSFARLEDVQGDE
                     ARLDVDQSQLLDPQDCLRVADPDLSVETDGQRDALCQSDPQRVFRAADCLDVCLRHCSA
                     PSSSPPSTPTPSSTTPTPSSSVSSVPTPSSPDSPSPTDHARSGDQHEVDLHDQRGDAA
                     GLRDQVDRSLGEDVRRLADGDGDQSERQQEIGHDEIRVGEGIEAVDFDKSAQGVEQGAF
                     LSQLQAVAQSDAADLQQVRDVDHDDVVAGFDGGRDHFGHDELDFFVTLPAHPVEGAI
                     SDRLRS"
     misc_feature    4460
                     /note="start codon according to author"
     CDS             4460. .5049
                     /note "Protein sequence is in conflict with the conceptual
                     translation"
                     /codon_start 1
                     /transl_table 11
                     /product "protease"
                     /protein_id "CAA89562.1"
                     /db_xref "GI:1333804"
                     /db_xref "SPTREMBL:Q52750"
                     /translation "SWNIRHEWWGLKmrqgGWRKGIdKGVIrcIrASqGSRISKGIKH
                     GSTWTNHSSTRRTGSCSRTLIPRWRTTGRGTPCASQTPSSSALTGLTSACGDVRPHP
                     ARRLHLLLRPAARPPPLAPQYLPYLHKARPTHRRLRTPPARCTSMRWTCTISAATRPV
                     SLIRWTIRWARDCASPTETESFSGSRSSAIISSGW"
BASE COUNT      1379 a    2439 c    2277 g    1219 t
ORIGIN


alignment scores:
            Quality:    7.00              Length:      7
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2 (1-4586) x AF000078.gb

 Align_seq 1/1  to reverse of: RFCGCP  from: 1  to: 7304

        3 ArgArgAspTrpSerProSer 9
          ||||||||||||||||||||||
     5012 CGGAGGGACTGGTCCTCCGTCT 4992

seq_name: qb_bal:AE000078

seq_documentation block:
LOCUS       AE000078    15586 bp    DNA             BCT       12-DEC-1997
DEFINITION  Rhizobium sp. NGR234 plasmid pNGR234a, section 1% of 4% of the
            complete plasmid sequence.
ACCESSION   AE000078 B00090
VERSION     AE000078.1  GI:2182439
KEYWORDS    .
SOURCE      Rhizobium sp. NGR234.
  ORGANISM  Plasmid Rhizobium sp. NGR234
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Rhizobiaceae; Rhizobium.
REFERENCE   1  (bases 1 to 15586)
  AUTHORS   Freiberg,C., Fellay,R., Bairoch,A., Broughton,W.J., Rosenthal,A.
            and Perret,X.
  TITLE     Molecular basis of symbiosis between Rhizobium and legumes
  JOURNAL   Nature 387 (6631), 394-401 (1997)
  MEDLINE   97305956
REFERENCE   2  (bases 1 to 15586)
  AUTHORS   Freiberg,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-1996) Genome Analysis, Institute for Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 15586)
  AUTHORS   Freiberg,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-1997) Genome Analysis, Institute for Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
  REMARK    Update by submitter
FEATURES             Location/Qualifiers
     source          1. .15586
                     /organism="Rhizobium sp. NGR234"
                     /plasmid "pNGR234a"
                     /strain "NGR234"
                     /db_xref "taxon:394"
     source          1. .15586
                     /organism="Rhizobium sp. NGR234"
                     /db_xref "taxon:394"
                     /clone "pXB182"
     misc_feature    complement(82. .130)
                     /note "nod box-like sequence"
                     /note "Region: nod box 10"
                     /evidence=not_experimental
     gene            complement(178. .870)
                     /gene "y4iC"
     CDS             complement(178. .870)
                     /gene "y4iC"
                     /note "hypothetical 25.8 kd protein"
                     /codon_start 1
                     /transl_table 11
                     /product "Y4iC"
                     /protein_id "AAB91698.1"
                     /db_xref "GI:2182440"
                     /translation "MYKERAFNAFARASGVDLLDAAAYGQQSNGDRRGFPRFGGDTER
                     LDCLQLGKVELLPGLIEVIGIAMRNLIFKSLDLLLLLELRNVLYVKRRVHMAGIRLSK
                     PASEIMIGQVIKALDGPLAPIHCASRSAFACDDCNDPVYRQSVLVICHRCKKLGYGVA
                     AWLKNAGRVELSDVSKRTIAQHQLKIGQRFRAHQYCPAGRENERNICIEMGGPLRFGII
                     VKLDRHIQSLVN"
     gene            complement(812. .2800)
                     /gene "y4iD"
     CDS             complement(812. .2800)
                     /gene "y4iD"
                     /note "probable monooxygenase"
                     /codon_start 1
                     /transl_table 11
                     /product "Y4iD"
                     /protein_id "AAB91699.1"
                     /db_xref "GI:2182441"
                     /translation "MAHSSDLALFRHKIAEALPSANVPILGLILYQFLGKEYWLIFPF
                     LPVKSVWGDNDSGGLAAELQTEIRNAALDAIICWHSGAPVIKQDLSEFFILRMLIVSE
                     AEPIPPEYADMMLHKLRRYAGAIPDPVSLPKGFRVLIIGAGMSGVAAARLRQLGISY
                     QVVEKQDSIGGVWHAHHYPGCGVDIPSHLYSYTFASGNWSTFFTQKFIIDDYFNRVAR
                     DFGIESSIRYGTECLVTRYDEESLTWHSRLRLPNGTEEILVTNVVLSAVGGFTTPKWP
                     NLSGLRNFDGPVVHTSKWDPEVALDGKRVAVIGNGASAMQVVPAIAGEVGALLIFQRS
                     QWVAPPFPKFQKPVPEPMQFLFRFVPHYWLYRLFLSWIYLSFVHGAIGKDPAWPHPD
                     RSVNAVNDRDREAYTQYIEGQLAGRPDLIAKVIPSYPPFGKRMLIDNGWYRTLLKPHV
```

```
                     ILVDGAAARVEGRSIHAIIDETHEAFLIVASGYFIPYIIEVEVIGNNSKIVRGVWI
                     DDDCQAYLGTVVAGFPNFFMLYGPNIALCHFCFIL...IDYVLEVLKQMTGKRIV
                     EVECRQHIVQHYNKKIQVMPQVMIWSHDCPTPNNI...YTNCPWPLVCTWNIEE
                     ADLDDYKIMPQVDSRLETSGVPREGVQRPGSKLRRRPS"
gene                 3544..3951
                     /gene="y41E"
CDS                  3544..3951
                     /gene="y41E"
                     /note="hypothetical 15.4 kd protein (gene 3',
                     fragmentous); similar to Y4rI,Y4zA"
                     /codon_start=1
                     /transl_table=11
                     /product="Y41E"
                     /protein_id="AAB91700.1"
                     /db_xref="GI:2182442"
                     /translation="MQEKLAALAGIIFAMGSIGVAQHIIFLASLRIAWKIAIKII
                     RPIVKAKRQRRPPDPIVFAIAPGSGPTAFIAVRFIGSIFVPLGVPFLIVPFSIQT
                     QKRLKVWRSEQADALLFGTLNKELPAQQIAKDH"
misc_feature         3544..3951
                     /gene="y41E"
                     /note="insertion sequence homology (ISH) 3a; region
                     containing putative IS element-derived ORF (y41E)"
                     /note="Region: ISBR1"
gene                 complement(4460..5377)
                     /gene="y41F"
CDS                  complement(4460..5377)
                     /gene="y41F"
                     /note="hypothetical 34.1 kd protein"
                     /codon_start=1
                     /transl_table=11
                     /product="Y41F"
                     /protein_id="AAB91701.1"
                     /db_xref="GI:2182443"
                     /translation="MSGAIIWAFAFAVPIFSSAIVFIINMVLAIFFLKFIGVSVY
                     VQWHHQDERCKKGISVLEVGPGIVFAPQQKTLPLVIGLIGFFIRCIIVFRSRDFL
                     AQLSSAILGLKIFFIPGDIFSGRRVFCDDPAAFIVMFCWFFSNIAVPMNGIHRMEA
                     TLKEISNSAKRCWVAISIGSDLKKGVAESYYEAIHFFLNVEYEMKAAELPVDGNFDPE
                     AFEYEFDVIGGDQFMQVIHIAVVFKNTGFTGKGIFISSDSSHHNSPFFSFSS
                     KCAHPAGFVPIPVGSLDAGSAIHVLQFVN"
misc_feature         5839..6158
                     /note="insertion sequence homology (ISH) 3b; region
                     containing putative IS element derived ORF (y41G)"
                     /note="Region: ISBR1"
gene                 complement(5839..6111)
                     /gene="y41G"
CDS                  complement(5839..6111)
                     /gene="y41G"
                     /note="hypothetical 10.5 kd protein (gene 3',
                     fragmentous); similar to Y4aF and Y4bF"
                     /codon_start=1
                     /transl_table=11
                     /product="Y41G"
                     /protein_id="AAB91702.1"
                     /db_xref="GI:2182444"
                     /translation="MLDLVRPHVIDFGGTLAAFFLIFPHPLAVSKETIPQWIMEAGIW
                     ALRKEMKKRVFLQPGRRPCFGELLQIGSPCIDRKSIVLITLSIAR"
gene                 6294..7013
                     /gene="y41H"
CDS                  6294..7013
                     /gene="y41H"
                     /note="hypothetical 26.8 kd protein (possibly fragment of
                     a larger protein originally encoded by y41K and y41L)"
                     /codon_start=1
                     /transl_table=11
                     /product="Y41H"
                     /protein_id="AAB91703.1"
                     /db_xref="GI:2182445"
                     /translation="MFTMVAMCIAAVIVAVALAIIAVGLNPSVSLIYSVPLLFVV
                     IPGDASHKLLKADFGYVDLNGVGWQINKIIKTDGASIPPLLQPFVGSPWFLGYIRAAV
                     IHDWYCDRHVRIFWKETRKVFYDIMLASGLEKPEAKLLFYAVYAPCPWCYLVPGEKCA
                     AGKNCIQMTGKDAAFVQLPGRLAPVSSAGELKATKATIDMLRRRSGDALTIDELMAIAD
                     EAHPKQTLKDQRPAGGDETIK"
gene                 complement(7186..9297)
                     /gene="y41I"
CDS                  complement(7186..9297)
                     /gene="y41I"
                     /note="hypothetical 76.2 kd integral membrane protein"
                     /codon_start=1
                     /transl_table=11
                     /product="Y41I"
                     /protein_id="AAB91704.1"
                     /db_xref="GI:2182446"
                     /translation="MKEKSTLCIVMIICIFLIICIAASLAIIGVLAQVEDELQ
                     AIDAWEKAKAEGAPDLFEYDVRLKAISGASAGALIAAIGVIAAGFEAIAIPSFAPG
                     SQNIRPTIGKLYRSWVISPTLVSPLGSPDLLSLEDLAGGKPVISVLNANVETATGAEA
                     LEAIGTLSPKAYVASSLHLYMMLSNLRGVPYATHFNGGQYNMMIHAEPVHYVVFGII
                     WKPTESPFALITSGTSTAATSLEGAAGASIKFLEWLAFANAALASGAFPIGLSPKVIA
                     IATSQYASAKFPISEPQSELRPTPIWPDAWHVSASQQIYPFSFVSVPFGLINNGPFFV
                     PFTLMKDPPPPNERNAEKADPAVIMIAPFPFCGCEPPLGVLSIAPPVVIALIGQ
                     GVFEKPIQLLAVAAEGTHSKFMISPHKVIPSTIPGQFEKFFLESTASIIGGQGFVLF
                     AFEIHFYQLGFFNQVFPLMFRLCTIFENHQTCIHWPFTAAFSENAVICFLSC SPMFYV
                     PIIPLVCIALIFFVPVPKWAKIDENAFALLVKGTEARLVAVAAKRLVSIFTISAFMKI
                     NFLLLVGRNRIVDYIRLTLIQGLVMRDQIFGWPLPAADLRFDPVKAVLAALCDPAFDL
                     SFPAIAKTCKIHGIVFFLSTLAVAAGANFQVWLAGWSFTLFLCFTLVSMGFSFL
                     AIILQGKSFAFLFAKPVVDKK"
source               7567..15586
                     /organism="Rhizobium sp. NGR234"
                     /db_xref="taxon:394"
                     /clone="pXB1661"
gene                 complement(9846..11136)
                     /gene="y41J"
CDS                  complement(9846..11136)
                     /gene="y41J"
                     /note="hypothetical 45.7 kd protein"
                     /codon_start=1
                     /transl_table=11
                     /product="Y41J"
                     /protein_id="AAB91705.1"
                     /db_xref="GI:2182447"
                     /translation="MKEKSNFWFLLLCFIVLVAATLYYVVFMFSVLLIVYFEVIEVIW
                     LEQNWSQSQRGWMHHADQGTVTFSMPYFWILAALLAPIFLTLTAGPPFLSSDYLDREGFI
                     IAFSSGLFVGLAHSGPLVFPFLIAQFWVNFAIGRPLLIVLLVAAVTGQAPLKAAVAVLE
                     VGGPALLIAGCFEPFASGLALLFIRYAFLSLPRFAIAVSIGQAPLKAAVAVLERQLIDV
                     LAGGKIEVDLDKKVAEKSIEEGFGRLDALNKIGNQVFSLLLERPENYVAQSAIVAFPH
                     IWGISWGIWVQIVNASLMQPMVFNAGFALGVPAFINLIKSEQPLFASIVSVDLIFFIFQ
                     GLAGKQPLALNGFIGLRFPFWFSNLFGSIGDKLALIGGAAVYADSQGVDLFFVGSFGF
                     WFQKHWTNENSAGIEVIFVFLINVENLCIGFAGAGPMAIPSFVELPSELGIPLDSFGSA
                     LGALVAKIAAKWYDNQLPPVPAEQREIMNGNRQNGIQAPLAYKAKFLCIWAIPFFLH
                     NSVFLIPALLSPAGLFFPLFWLGNREYFYPPELLYLLLLSGGFFFCSRFPGNSNAGH
                     EFSDFPGPGVIGPALKPDEKAALIAYLKTL"
gene                 complement(11189..11887)
                     /gene="y41K"
CDS                  complement(11189..11887)
                     /gene="y41K"
                     /note="hypothetical 26.8 kd protein (possibly fragment of
                     a larger protein originally encoded by y41K and y41L)"
                     /codon_start=1

alignment_scores:
                 quality:    7.00              Length:       7
                   Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 [1] [2] [3] [4] [5] [6] [7] [8] [9] [10] AE000078/rev [1]

 Align set 1/1  to reverse of: AE000078  from: 1  to: 15586

        5 ArgAlaAspIlePheSerProSer 9
          |||||||||||||||||||||
    7567 ACGGCGATTTCAGCCCGTTCG 7653

seq_name: sb:ba:sC035

seq_documentation_block:
LOCUS       SC035      36394 bp    DNA             BCT       13 MAR 2000
DEFINITION  Streptomyces coelicolor cosmid D35.
```

```
ACCESSION   AL160312
VERSION     AL160312.1  GI:7242710
KEYWORDS    3-oxoacyl-[acyl carrier protein] reductase; acetyltransferase; cell
            division related protein; cold shock protein; dehydratase; DNA
            polymerase related protein; integral membrane protein; ketoacyl CoA
            thiolase; long-chain fatty acid CoA ligase; methyltransferase;
            protein kinase; proton transport protein; secreted protein;
            serine/threonine protein kinase; tetR family regulatory protein.
SOURCE      Streptomyces coelicolor A3(2).
  ORGANISM  Streptomyces coelicolor A3(2)
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 36394)
  AUTHORS   Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J.,
            Kinashi,H. and Hopwood,D.A.
  TITLE     A set of ordered cosmids and a detailed genetic and physical map
            for the 8 Mb Streptomyces coelicolor A3(2) chromosome
  JOURNAL   Mol. Microbiol. 21 (1), 77-96 (1996)
  MEDLINE   97030451
REFERENCE   2  (bases 1 to 36394)
  AUTHORS   Sanger,K.J. and Barris,D.
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 36394)
  AUTHORS   Cerdeno,A.M., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAR-2000) Streptomyces coelicolor sequencing project,
            Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA E-mail: barrell@Sanger.ac.uk Cosmids supplied by Prof.
            David A. Hopwood, J.C. John. Innes Centre, Norwich Research Park,
            Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT     Notes:
            Streptomyces coelicolor sequencing at the Sanger Centre is funded
            by the BBSRC and Beowulf Genomics
            Details of S. coelicolor sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL: http://www.sanger.ac.uk/Projects/S_coelicolor/)
            CDS are numbered using the following system eg SCJB7.01c, SC (S.
            coelicolor), JB7 (cosmid name), .01 (first CDS), c (complementary
            strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous.
            The length in codons is given for each CDS.
            Usually the highest scoring match found by tfasta - o is given for
            CDS which show significant similarity to other CDS in the database.
            The position of possible ribosome binding site sequences are given
            where these have been used to deduce the initiation codon.
            Gene prediction is based on positional base preference in codons
            using a specially developed Hidden Markov Model (Krogh et al.,
            Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot
            program of Bibb et al., Gene 40:157-66(1984) as implemented at
            http://www.nih.go.jp/
            jun/cgi-bin/frameplot.pl CAUTION: We may not have predicted the
            correct initiation codon. Where possible we choose an initiation
            codon (atg, gtg, ttg or (att)) which is preceded by an upstream
            ribosome binding site sequence (optimally 5-13bp before the
            initiation codon). If this cannot be identified we choose the most
            upstream initiation codon.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
            sequenced clone. It may be shorter because we only sequence
            overlapping sections once, or longer, because we arrange for a
            small overlap between neighbouring submissions.
            Cosmid D35 .
FEATURES             Location/Qualifiers
     source          1..36394
                     /organism="Streptomyces coelicolor A3(2)"
                     /strain="A3(2)"
                     /db_xref="taxon:100226"
                     /clone="cosmid D35"
     misc_feature    1..95
                     /gene="SCD35.01"
                     /note="nominal overlap with S. coelicolor cosmid StD69"
     gene            1..181
                     /gene="SCD35.01"
     CDS             1..181
                     /partial
                     /gene="SCD35.01"
                     /note="Protein sequence is in conflict with the conceptual
                     translation; SCD35.01, unknown, len: 95 aa"
                     /codon_start=2
                     /transl_table=11
                     /product="hypothetical protein SCD35.01"
                     /protein_id="CAB77285.1"
                     /db_xref="GI:7242711"
                     /translation="MTRDAMRPGTPNELTREELLELRSAVTRPVEENTRYETLREYE
                     GPDPARRESPQPMRV"
     gene            406..1044
                     /gene="SCD35.02"
     CDS             406..1044
                     /gene="SCD35.02"
                     /note="SCD35.02, possible DNA polymerase related protein;
                     len: 212 aa; similar to TR:AAF11304 (EMBL:AE002016)
                     Deinococcus radiodurans DNA polymerase related protein
                     DR1751, 237 aa; fasta scores: opt: 880 z-score: 451.4 E():
                     1e-1; 41.5% identity in 188 aa overlap and to N-terminal
                     region of SW:DPOL_BPSP1 (EMBL:M84415) Bacteriophage SP01
                     DNA polymerase (EC 2.7.7.7) 31, 924 aa; fasta scores: opt:
                     191 z-score: 223.4 E(): 5.2e-05; 33.7% identity in 169 aa
                     overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="putative DNA polymerase related protein"
                     /protein_id="CAB77286.1"
                     /db_xref="GI:7242712"
                     /translation="MAIEDDYIAQFEVFDKGLLALKAAAAECRQGFEDRDAIQIVEG
                     AGKASAKVMLVGEQPGGQEDRQGKPFVGPAGRLLDKALAEAGLDPADAYVTNAVKHFK
                     PIRAEPRKRRIHKAPTLRETAACGPWLAAELDRVEPELIVVLGATAGKALLGSSFRVT
                     RVRGTVLEEETHGRPQRLVPTVHPSAVLRADDREAAYRGLLSDLEVAARALA"
     gene            1061..1585
                     /gene="SCD35.03"
     CDS             1061..1585
                     /gene="SCD35.03"
                     /note="SCD35.03, possible acetyltransferase, len: 174 aa;
                     similar to TR:CAB71250 (EMBL:AL138598) Streptomyces
                     coelicolor putative acetyltransferase SCIA2.12, 181 aa;
                     fasta scores: opt: 214 z-score: 278.1 E(): 4.6e-08; 42.9%
                     identity in 149 aa overlap and to SW:ITIC_PSESY
                     (EMBL:X17150) Pseudomonas syringae (pv. syringae)
                     acetyltransferase (EC 2.3.1.-) 177, 17 aa; fasta scores:
                     opt: 137 z-score: 182.0 E(): 0.01; 28.8% identity in 139
                     aa overlap. Contains Pfam match to entry PF00583
                     Acetyltransf, Acetyltransferase (GNAT) family"
                     /codon_start=1
                     /transl_table=11
                     /product="putative acetyltransferase"
                     /protein_id="CAB77287.1"
                     /db_xref="GI:7242713"
                     /translation="MSLTFTEDFAVIPDLCRGVIDLWTDVSNAGSSVGLVPEVIREDV
                     RPELLKHLIGMAEGRTRLLVGHDEEGRVAAFAFLALNTHRLMRHWLWLYIVMVHPEHQ
                     GRGYGRDLMTAAAEAARTCDGIEAIRLTCRGDLGLERFYESCGYKEVGFVPGATKVAF
                     GDDRDDVPMLLPLA"
     misc_feature    1082..1501
                     /gene="SCD35.03"
                     /note="Pfam match to entry PF00583 Acetyltransf,
                     Acetyltransferase (GNAT) family, score 33.50, E-value
                     4.9e-06"
     gene            1706..2005
                     /gene="SCD35.04"
     CDS             1706..2005
                     /gene="SCD35.04"
                     /note="SCD35.04, possible secreted protein, len: 297 aa.
                     Contains possible N-terminal region signal peptide
                     sequence"
                     /codon_start=1
                     /transl_table=11
                     /product="putative secreted protein"
                     /protein_id="CAB77288.1"
```

```
                     /DB_xref="GI:7242731"
                     /translation="MSVSSLSSPTKYLEVPTWAQTLALGLVLVCVLLGWLAEPQLTSWLV
                     TLLEKVGDTFIGLLKLAVAPLVFATLVSTLNELEVRNAARLASKTLLWFMTTSLIAV
                     AIGLVIGLVTNPGAGTGLTAADGAKPEHTGSWIDFLTGIVPTDVITPFTELNVLQIVF
                     MAAVACIAAIQICPKAQPLINLSESISEIFQKAIWWVMKLAPIGVFGLGKAIAYGW
                     DLLIGKYAFFTADIYVGCALVMFVVYPTLLAIVAKASPLPFFKGAWPAIQLAFVSRSSV
                     GTMPLTQKVTERLGVPKEYASFAVPFGATIKMDGCAAIYPAIAAIFVAQIFDIQLGVG
                     DYLLIAFVSVVGSAATAGLTGATVMLTLLISTLGLPMEGVGLLLAIDPFIDMMRTAIN
                     VAGQALVPVIVSARECLLDRKAYDEAHASPIDEPEREEQASEDVEVAA"
     misc_feature    2468..3537
                     /gene="SCD35.05"
                     /note "Pfam match to entry PF00375 SDF,
                     Sodium:dicarboxylate symporter family, score 287.90,
                     E-value 1.3e-82"
     gene            3703..4425
                     /gene "SCD35.06"
     CDS             3703..4425
                     /gene="SCD35.06"
                     /note "SCD35.06, probable tetR-family regulatory protein,
                     len: 240 aa; similar to TR:O54664 (EMBL:AL021929)
                     Mycobacterium tuberculosis putative regulatory protein
                     MTV034.04, 204 aa; fast scores: opt: 447 z-score: 567.8
                     E(): 7.5e-21; 41.4% identity in 198 aa overlap and to
                     SW:MTRR_NEIGO (EMBL:225797) Neisseria gonorrhoeae
                     regulatory protien MtrR, 210 aa; fasta scores: opt: 164
                     z-score: 193.8 E(): 0.0024; 28.3% identity in 159 aa
                     overlap. Contains Pfam match to entry PF00440 tetR,
                     Bacterial regulatory proteins, tetR family and possible
                     helix-turn-heli[illegible]"

alignment_scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_38_46 x SCD35  ..

 Align Seq 1/1  to: SCD35  from: 1  to: 46364

       3 ArgArgAspIrpSerProSer 9
         ||||||||||||||||||||
    9179 CGGCGGCACTGGCCAGCGGC 9196

seq_name: qb_ba3:SC1C2

seq_documentation_block:
LOCUS       SC1C2       42210 bp    DNA             BCT       15-JAN-1999
DEFINITION  Streptomyces coelicolor cosmid 1C2.
ACCESSION   AL031124
VERSION     AL031124.1  GI:3455667
KEYWORDS    [illegible]; 3-isopropylmalate dehydratase large subunit; 3-isopropylmalate
            dehydratase small subunit; 3-isopropylmalate dehydrogenase;
            branched-chain amino-acid aminotransferase; carboxyl transferase;
            delta 1 pyrroline 5 carboxylate dehydrogenase; glnX; glutamyl tRNA
            synthetase; histone-like DNA binding protein; hydrolase; ilvE;
            leuB; leuC; leuD; lyase; secreted lyase; transfer RNA Glu;
            transfer RNA Glu; ureAB; urease alpha subunit; urease beta and
            gamma subunits; ureC.
SOURCE      Streptomyces coelicolor A3(2).
  ORGANISM  Streptomyces coelicolor A3(2)
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 42210)
  AUTHORS   Murphy,L. and Harris,D.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 42210)
  AUTHORS   Barrell,B.G. and Rajandream,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1998) Streptomyces coelicolor sequencing project,
            Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA Email: barrell@sanger.ac.uk Cosmids supplied by Prof.
            David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
            Colney, Norwich, Norfolk NR4 7UH, UK
REFERENCE   3  (bases 1 to 42210)
  AUTHORS   Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J.,
            Kinashi,H. and Hopwood,D.A.
  TITLE     A set of ordered cosmids and a detailed genetic and physical map
            for the 8 Mb Streptomyces coelicolor A3(2) chromosome
  JOURNAL   Mol. Microbiol. 21 (1), 77-96 (1996)
  MEDLINE   97000351
COMMENT     Notes:
            Streptomyces coelicolor sequencing at the Sanger Centre is funded
            by the BBSRC.
            Details of S. coelicolor sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL: http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are
            numbered using the following system eg SCD7.01c: SC (S.
            coelicolor), D7 (cosmid name), .01 (first CDS), c (complementary
            strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous. The length
            in codons is given for each CDS.
            Usually the highest scoring match found by fasta -o is given for
            CDS which show significant similarity to other CDS in the database,
            the position of possible ribosome binding site sequences are given
            where these have been used to deduce the initiation codon. Gene
            prediction is based on positional base preference in codons using a
            specially developed Hidden Markov Model (Krogh et al., Nucleic
            Acids Research, 22(22):4768-4778(1994)) and the FrameFlot program
            of Bibb et al., Gene 30:157-66(1984) as implemented at
            http://www.[illegible].ac.uk/Users/jun/c++_bin/framefloat.pl. CAUTION: We may not have predicted the
            correct initiation codon. Where possible we choose an initiation
            codon (atg, gtg, ttg or (att)) which is preceded by an upstream
            ribosome binding site sequence (optimally 5-13bp before the
            initiation codon). If this cannot be identified we choose the most
            upstream initiation codon.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
            sequenced clone. It may be shorter because we only sequence
            overlapping sections once, or longer, because we arrange 1 or a
            small overlap between neighbouring submissions. Cosmid 1C2 lies
            between 8D9 and 7A1 on the AseI-B genomic restriction fragment.
FEATURES             Location/Qualifiers
     source          1..42210
                     /organism="Streptomyces coelicolor A3(2)"
                     /strain "A3(2)"
                     /db_xref "taxon:100226"
                     /clone "cosmid 1C2"
     gene            1..1228
                     /gene "SC1C2.01"
     CDS             1..1228
                     /gene "SC1C2.01"
```

```
                     /note "SCIC2.01, probable delta 1 pyrroline 5 carboxylate
                     dehydrogenase, partial CDS, len: 408 aa; similar to many
                     e.g. TR:O50443 (EMBL:AL010186)
                     delta 1 pyrroline 5 carboxylate dehydrogenase (M.
                     tuberculosis) (543 aa), fasta scores; opt: 1753 z score:
                     2214.2 E(): 0, 65.1% identity in 407 aa overlap, and
                     P5C2_HUMAN delta 1 pyrroline 5 carboxylate dehydrogenase
                     (563 aa), fasta scores; opt: 1273 z score: 1730.4 E(): 0,
                     48.5% identity in 408 aa overlap. Contains PS00687 and
                     PS00070 Aldehyde dehydrogenases glutamic acid active site
                     and cysteine active site and Pfam match to entry PF00171
                     aldedh, Aldehyde dehydrogenases, score 114.70, E value
                     1.8e-30"
                     /codon_start=2
                     /transl_table=11
                     /product="delta 1 pyrroline 5 carboxylate dehydrogenase"
                     /protein_id="CAA19968.1"
                     /db_xref="GI:3355668"
                     /db_xref="SPTREMBL:O86502"
                     /translation="IDSPCELIDFWRENVHYARNILAEQPPANSQGVWNRMDHRPLEG
                     FVYAITPFNFSAIAANLPTAPALMGNVVVWKPSPTQITHAAVLLMQLLEEAGLPEGVIN
                     LVIGDGIAVSEVALEHRDLAGIHFTGSTKIFQHLWKTVGNNIEKYRTYPRLVGETGGK
                     DFLVAHPSADRAVLKTALIMGAFEYQGQKCSAISRAYIPASIWNIMKFKEEFAAEVPAL
                     IMGDVTDLSNFIGAVIDERSFAKNKAAIDKAKEDEICLIVAGGSYDDSVGYFVRPTVV
                     ECTDPRNEVFRTEYFGPFLAVHVYDDSADDAYDAMLTQMESVSDYALTGSVISNDRAA
                     AAYIMEFIEYAAGNIYINDKSTGAVVGQQPFSGARCSGTNDKAGSPQYMMRWTSPQA
                     KEILVAPTDYIYPHMG"
misc_feature         1..3807
                     /note="true overlap with cosmid MD9"
misc_feature         2..1189
                     /gene="SCIC2.01"
                     /note="Pfam match to entry PF00171 aldedh, Aldehyde
                     dehydrogenases, score 114.70, E value 1.8e-30"
misc_feature         464..487
                     /gene="SCIC2.01"
                     /note="PS00687 Aldehyde dehydrogenases glutamic acid
                     active site"
misc_feature         548..583
                     /gene="SCIC2.01"
                     /note="PS00070 Aldehyde dehydrogenases cysteine active
                     site"
stem_loop            1246..1283
                     /note="hairpin loop with 18bp stem"
RBS                  1522..1527
                     /note="possible RBS upstream of SCIC2.02"
gene                 1547..1668
                     /gene="SCIC2.02"
CDS                  1547..1668
                     /gene="SCIC2.02"
                     /note="SCIC2.02, questionable ORF, len: 43 aa"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein SCIC2.02"
                     /protein_id="CAA19969.1"
                     /db_xref="GI:3355669"
                     /db_xref="SPTREMBL:O86503"
                     /translation="MAETTVRRRVRHLCFTTDCDRKNAAAALQRCLDRRDNGGATGH"
gene                 1744..2787
                     /gene="leuA"
CDS                  1744..2787
                     /gene="leuA"
                     /note="SCIC2.03, leuB, probable 3 isopropylmalate
                     dehydrogenase, len: 347 aa; similar to e.g. LEU3_CORGL
                     3 isopropylmalate dehydrogenase (EC 1.1.1.85) (340 aa),
                     fasta scores; opt: 1450 z score: 1418.1 E(): 0, 68.0%
                     identity in 338 aa overlap. Contains PS00470 Isocitrate
                     and isopropylmalate dehydrogenases signature and Pfam
                     match to entry PF00180 isodh, Isocitrate and
                     isopropylmalate dehydrogenases, score 356.60, E value
                     1.4e-105"
                     /codon_start=1
                     /transl_table=11
                     /product="3 isopropylmalate dehydrogenase"
                     /protein_id="CAA19970.1"
                     /db_xref="GI:3355670"
                     /db_xref="SPTREMBL:O86504"
                     /translation="MSESLNLAVIGLTLGAVVAICLFVLCAVGLSVFELGKFLDF
                     GARKYHATGELLTDADLDALKAHDATLGATGDPSVPSGVLEKGFLLKLRFAPGHHVN
                     LKPSKLLPGVATPLAGQPEIDFVVVRENTEGFYIGNGGTIKKGTEHEVAIEVSVNTAY
                     SVERVVRFAFAFAQARPRKKLTLVHKNNVLTFAGHEWRNTFNEVAAEYPEVTLDYLHV
                     DAATLELVTDPARFDVIVTDNLFGDILSDLAAAVAGSIGVAASGNINPSGDFPSMFEP
                     VHGSAPDIAGQGKADPTATVLSVALLLRHLGYEDEAARIEDAVSADLCERGILPARST
                     SEIGDLLAARVAG"
misc_feature         1759..2766
                     /gene="leuB"
                     /note="Pfam match to entry PF00180 isodh, Isocitrate and
                     isopropylmalate dehydrogenases, score 356.60, E value
                     1.4e-105"
misc_feature         2458..2517
                     /gene="leuB"
                     /note="PS00470 Isocitrate and isopropylmalate
                     dehydrogenases signature"
RBS                  3038..3042
                     /note="possible RBS upstream of ilvE"
gene                 3052..4140
                     /gene="ilvE"
CDS                  3052..4140
                     /gene="ilvE"
                     /note="SCIC2.04, ilvE, probable branched chain amino acid
                     aminotransferase, len: 362 aa; similar to many, e.g.
                     ILVE_BACSU putative branched chain amino acid
                     aminotransferase (362 aa), fasta scores; opt: 1047
                     z score: 1401.5 E(): 0, 44.7% identity in 360 aa overlap.
                     Contains PS00770 Aminotransferases class IV signature"
                     /codon_start=1
                     /transl_table=11
                     /product="branched chain amino acid aminotransferase"
                     /protein_id="CAA19971.1"
                     /db_xref="GI:3355671"
                     /db_xref="SPTREMBL:O86505"
                     /translation="MTTPTIELKPSAHPLSDSFRAATLANHGFGKHFTDHMVTIKWTE
                     GKGWHQQIVPYAPISIDPATMVLHYAQEIFEGLKAYRRPDGSVALFRPDKNGARFQA
                     SSPPICMPELPVDVFIEACDALVAQDEKWVPAHGGEESLYLRPFMIATEVGLGVKPAN
                     EYLFTVIASPAGAYFPGGVKPVSIWVSEDRVRAVPGGMGDAKTGGNYAASLLAQAEAA
                     AKGYQQVCYLDAIERKWVEEIGGMNLYFVYGNKIVTPSLTGSILEGVTRDSLITVARD
                     LGYEAEERGRVSVDQWQRDSENGTLTEVFACGTAAVITPVGTVKRAGAQWQQSGGEIGE
                     VTQRLRDALLDIQRGTVADPHGWMHTLA"
misc_feature         3763..4852
                     /gene="ilvE"
                     /note="PS00770 Aminotransferases class IV signature"
gene                 complement(4214..5644)
                     /gene="SCIC2.05c"
CDS                  complement(4214..5644)
                     /gene="SCIC2.05c"

alignment_scores:
                Quality:    7.00              Length:       7
                  Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 DS:09:126:945b-2_copy_38_46.x_SCIC2_rev  ..

 Align Seq 1/1  to reverse of: SCIC2  from: 1  to: 42210

        5 ArgAlaAspIleSerProSer 9
          ||||||||||||||||||||
    20919 AGGCGGACTGSTGCCATCC 20899
```

OM nt: US-09-126-945B-2_COPY_38_46 [illegible]

Date: Sep 14 2001 [illegible] PM

About: Results were produced by [illegible]
Copyright (c) [illegible]

Command line parameters:
MODEL frame+_p2n.model -DEV xlp
[illegible]
-DB=N_Geneseq_0601 -QFMT fastap -SUFFIX=olip2n.rng -GAPOP 4.500
GAPEXT 0.050 -MINMATCH 0.100 -LOOPCL 0.000 -LOOPEXT 0.000
-QGAPOP 4.500 -QGAPEXT 0.050 -XGAPOP 60.000 -XGAPEXT 60.000
-FGAPOP 6.000 -FGAPEXT 7.000 -YGAPOP 60.000 -YGAPEXT 60.000
-DELOP 6.000 -DELEXT 7.000 -START 1 -MATRIX [illegible]
-TRANS human40.cdi -LIST 45 -DOCALIGN 200 -THR_SCORE quality
-THR_MIN 1 -ALIGN 15 -MODE LOCAL -OUTFMT pto -NORM ext -MINLEN 0
-MAXLEN 2000000000 -USER US09126945_@CGN1_1_308 -NCPU 6 -ICPU 3
-LONGLOG -NO_XLPXY -WAIT -THREADS 1

Search information block:
Query: US-09-126-945B-2_COPY_38_46
Query length: 9
Database: N_Geneseq_0601:*
Database sequences: 730101
Database length: 313950809
Search time (sec): 680.150000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
/SIDS1/gsqdata/geneseq/geneseqp/NA2000.DAT:AAZ50691	+	9.00	163.01	0.4145	1894	!
/SIDS1/gsqdata/geneseq/geneseqp/NA2001.DAT:AAC83261	+	9.00	163.01	0.4145	1894	!
[illegible] NA2001.DAT:AA[illegible]	+	[illegible]	155.27	[illegible]	3317	!
[illegible] NA1998.DAT:AAV[illegible]		7.[illegible]	87.75	[illegible]	[illegible]	!
[illegible] NA1998.DAT:AAV[illegible]		7.[illegible]	87.73	[illegible]	536165	!
[illegible] NA2000.DAT:AAA[illegible]		[illegible]	[illegible]	20.29	38	!
[illegible] NA1998.DAT:AAX[illegible]		6.[illegible]	120.38	102.[illegible]	[illegible]	!
[illegible] NA2000.DAT:AAC04518		6.[illegible]	119.6[illegible]	122.61	[illegible]	!
[illegible] NA1992.DAT:AAQ[illegible]	+	6.00	119.25	128.96	[illegible]	!
[illegible] NA2001.DAT:AA[illegible]		[illegible]	117.21	147.[illegible]	386	!
[illegible] NA2001.DAT:AA[illegible]		6.[illegible]	111.28	211.[illegible]	[illegible]	!
[illegible] NA1999.DAT:AAT12247		6.[illegible]	[illegible]	219.78	71[illegible]	!
[illegible] NA2000.DAT:AAA[illegible]		6.[illegible]	[illegible]	[illegible]	[illegible]	!
[illegible] NA1995.DAT:AAT[illegible]		6.[illegible]	111.71	257.12	676	!
[illegible] NA2000.DAT:AAA[illegible]		[illegible]	111.71	257.12	676	!
[illegible] NA2001.DAT:AAF7[illegible]		[illegible]	[illegible]	[illegible]	[illegible]	!
[illegible] NA2001.DAT:AAF7[illegible]		[illegible]	[illegible]	327.[illegible]	[illegible]	!
[illegible] NA2000.DAT:AA[illegible]		[illegible]	[illegible]	[illegible]	[illegible]	!
[illegible] DAT:AA[illegible]	+	6.00	106.91	371.57	1147	!
[illegible] NA2001.DAT:AAF8[illegible]	+	[illegible]	105.82	[illegible]	1167	!
[illegible] NA2000.DAT:AAC[illegible]	+	[illegible]	106.26	408.69	1270	!
[illegible] NA2000.DAT:AAC75866	+	6.00	106.02	419.51	2065	!
[illegible] NA1997.DAT:AAT84487	+	[illegible]	118.85	115.41	19	!
[illegible] NA2000.DAT:AAA72069	+	[illegible]	[illegible]	152.74	21	!
[illegible] NA1997.DAT:AAT88477	+	5.00	116.40	163.57	26	!

[illegible] NA1996.DAT:AAT[illegible] [illegible]
[illegible] NA1994.DAT:AAZ73160 + 5.00 115.[illegible] 195.37
[illegible] NA2000.DAT:AAD00981 + [illegible] 114.[illegible] 209.58
[illegible] NA2000.DAT:AAA[illegible]276 5.00 114.0[illegible] 221.21
[illegible] NA1994.DAT:AAZ73159 + [illegible]
[illegible] NA1999.DAT:AAZ2[illegible] [illegible] 11[illegible].87 226.32

seq_name: [illegible] NA2000.DAT:AAZ50691

seq_documentation_block:
ID AAZ50691 standard; cDNA; 1894 BP.
XX
AC AAZ50691;
XX
DT 23-MAY-2000 (first entry)
XX
DE Human prostate derived Ets factor cDNA.
XX
KW Prostate Derived Ets factor; PDEF; human; chromosome 6p21.3; cancer;
KW loss of heterozygosity; chromosomal translocation; linkage analysis;
KW cytostatic; cardiant; immunosuppressive; cerebroprotective; fungicide;
KW antibacterial; vulnerary; neuroprotective; antiparkinsonian; nootropic;
KW anabolic; antiinflammatory; anorectic; hybridisation probe; forensic;
KW tumour marker; diagnosis; treatment; prostate cancer; blood coagulation;
KW autoimmune disorder; haematopoietic; immune/nervous system; stroke;
KW neoplasm; microbial infection; tissue regeneration; heart attack;
KW scarring; food additive; preservative; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 416..1423
FT /*tag= a
FT /product= "Human prostate derived Ets factor"
FT /note= "Expressed in prostate epithelium"
XX
PN WO200006[illegible]-A1.
XX
PD 10-FEB-2000.
XX
PF 02-AUG-1999; 99WO-US17470.
XX
PR [illegible]-1998; 98US-0126945.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI [illegible] TA, [illegible], Fawzeh CA, Pedress CA, Rosen CA;
XX
DR WPI; [illegible].
DR P-PSDB; AAY45003.
XX
PT Novel prostate derived polypeptide, polynucleotide useful for
PT diagnosis, prevention and treatment of prostate cancer, autoimmune
PT disorders, microbial infections and also as food additive or
PT preservative -
XX
PS Claim 4; Fig 1; 142pp; English.
XX
CC The present cDNA sequence encodes a human prostate derived Ets factor
CC (PDEF), isolated from a human prostate carcinoma cDNA library
CC (ATCC No [illegible]). PDEF gene expression is highest in tissues involved in
CC endocrine [illegible], with highest expression in prostate epithelium
CC cells. PDEF gene is mapped to the human chromosome 6p21.3
CC region that is associated with loss of heterozygosity and chromosomal
CC translocations in various human cancers. PDEF has cytostatic, cardiant,
CC immunosuppressive, cerebroprotective, fungicide, antibacterial,
CC vulnerary, neuroprotective, antiparkinsonian, nootropic, anabolic,
CC antiinflammatory and anorectic activity. PDEF polynucleotides are useful
CC in linkage analysis as markers, as hybridisation probes for differential
CC identification of the tissues or cell types and as polymorphic markers
CC for forensic purposes. PDEF is useful as prostate specific tumour marker
CC for the diagnosis and treatment of prostate cancer. PDEF sequences are

CC useful for treating autoimmune disorders, haematopoietic, blood
CC coagulation, immune and nervous system disorders, hyperproliferative
CC disorders like, neoplasms and microbial infections, heart attacks,
CC stroke, scarring and for tissue regeneration. They are also useful as
CC food additives or preservatives.
XX
SQ Sequence 1894 BP; 368 A; 653 C; 571 G; 302 T; 0 other;

alignment scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2-COPY_38_46 x AAZ50691 .

Align seq 1/1 to: AAZ50691 from: 1 to: 1894

1 LeuGluArgArgAspTrpSerProSer 9
|||||||||||||||||||||||||||
527 CTCGAGACGGGACTGGAGTCCCAGT 553

seq_name: /cf1-1/1_data/pcore/seq_pcore/seq_NA/1_DAT/AAZ50691

seq documentation block:
ID AAC83261 standard; DNA; 1894 BP.
XX
AC AAC83261;
XX
DT 16-MAR-2001 (first entry)
XX
DE Gland specific Ets transcription factor (GSEF) nucleotide sequence.
XX
KW Transcription factor; gland specific Ets transcription factor; GSEF;
KW metastatic potential; cancer; tumour; metastasis; breast; prostate;
KW leukaemia; lymphoma; sarcoma; melanoma; chromosome 6p21.1-6p21.4; ds.
XX
OS Homo sapiens.
XX
PN WO200070092-A1.
XX
PD 23-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US13174.
XX
PR 14-MAY-1999; 99US-0134112.
XX
PA (CHIR-) CHIRON CORP.
XX
PI Kaufmann J, Xin H, Barrowe G;
XX
DR WPI; 2001-041019/05.
DR P-PSDB; AAB49628.
XX
PT Detecting metastatic and potential metastatic cancerous cells, useful
PT for diagnosing, prognosing, grading and staging of cancers by detecting
PT gland specific Ets transcription factor gene product in a biological
PT sample from a cell -
XX
PS Disclosure; Page 81-84; 95pp; English.
XX
CC This invention relates to a method for the detection and determination of
CC the metastatic potential of a cell. The method comprises detecting a
CC gland specific Ets transcription factor (GSEF) gene product in a test
CC sample. Detection of a GSEF gene product in the test sample in amount
CC lower than that in a normal cell, is indicative of a cell with high
CC metastatic potential. The method is useful for determining the metastatic
CC potential of a cell, for the diagnosis and prognosis of cancer as well as
CC grading and staging of cancers by detecting GSEF expression in a
CC biological test sample. The method may also be used to monitor patients
CC having a predisposition to develop a particular cancer. GSEF polypeptides
CC are useful for producing antibodies, in cancer diagnosis, prognosis,
CC grading, staging and management of breast and prostate tumours, and in
CC detecting polymorphisms in the sequence. GSEF genes and proteins are also
CC useful in gene therapy. GSEF gene product expression levels can be used
CC in conjunction with any tissue in which an alteration in GSEF gene
CC product expression levels is associated with development of a
CC cancer associated phenotype. Cancers, which can be monitored include
CC cancers of the prostate, cervix, lung and colon, melanomas, colorectal
CC adenocarcinomas, Wilm's tumour, retinoblastoma, sarcomas, myosarcomas,
CC lung carcinomas, leukaemia, and lymphomas. The GSEF gene is located on
CC human chromosome 6, specifically at 6p21.1-6p21.4. The present sequence
CC represents the DNA encoding GSEF.
XX
SQ Sequence 1894 BP; 368 A; 653 C; 571 G; 302 T; 0 other;

alignment scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2-COPY_38_46 x AAC83261 ..

Align seq 1/1 to: AAC83261 from: 1 to: 1894

1 LeuGluArgArgAspTrpSerProSer 9
|||||||||||||||||||||||||||
527 CTCGAGACGGGACTGGAGTCCCAGT 553

seq_name: /cf1-1/1_data/pcore/seq_pcore/seq_NA2001/DAT/AAC83266

seq documentation block:
ID AAC83266 standard; DNA; 3817 BP.
XX
AC AAC83266;
XX
DT 16-MAR-2001 (first entry)
XX
DE Gland specific Ets transcription factor (GSEF) cDNA sequence.
XX
KW Transcription factor; gland specific Ets transcription factor; GSEF;
KW metastatic potential; cancer; tumour; metastasis; breast; prostate;
KW leukaemia; lymphoma; sarcoma; melanoma; chromosome 6p21.1-6p21.4; ss.
XX
OS Homo sapiens.
XX
PN WO200070092-A1.
XX
PD 23-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US13174.
XX
PR 14-MAY-1999; 99US-0134112.
XX
PA (CHIR-) CHIRON CORP.
XX
PI Kaufmann J, Xin H, Barrowe G;
XX
DR WPI; 2001-041019/05.
DR P-PSDB; AAB49633.
XX
PT Detecting metastatic and potential metastatic cancerous cells, useful
PT for diagnosing, prognosing, grading and staging of cancers by detecting
PT gland specific Ets transcription factor gene product in a biological
PT sample from a cell
XX
PS Disclosure; Fig 2; 95pp; English.
XX
CC This invention relates to a method for the detection and determination of
CC the metastatic potential of a cell. The method comprises detecting a
CC gland specific Ets transcription factor (GSEF) gene product in a test
CC sample. Detection of a GSEF gene product in the test sample in amount
CC lower than that in a normal cell, is indicative of a cell with high

```
CC   metastatic potential. The method is useful for determining the metastatic
CC   potential of a cell, for the diagnosis and prognosis of cancer as well as
CC   grading and staging of cancers by detecting GSEF expression in a
CC   biological test sample. The method may also be used to monitor patients
CC   having a predisposition to develop a particular cancer. GSEF polypeptides
CC   are useful for producing antibodies, in cancer diagnosis, prognosis,
CC   grading, staging and management of breast and prostate tumours, and in
CC   detecting polymorphisms in the sequence. GSEF genes and proteins are also
CC   useful in gene therapy. GSEF gene product is expressed in high levels in
CC   in conjunction with any tissue in which an alteration in GSEF gene
CC   product expression levels is associated with development of a
CC   cancer-associated phenotype. Cancers which can be monitored include
CC   cancers of the prostate, cervix, lung and colon, melanomas, colorectal
CC   adenocarcinomas, Wilm's tumour, retinoblastoma, sarcomas, myosarcomas,
CC   lung carcinomas, leukaemia, and lymphomas. The GSEF gene is located on
CC   human chromosome 6, specifically at 6p21.1-6p21.3. The present sequence
CC   represents the cDNA encoding GSEF.
XX
SQ   Sequence 3417 BP; 719 A; 1026 C; 979 G; 693 T; 0 other;


Alignment_scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_38_46 x AAC83266   ..

 Align seq 1/1  to: AAC83266  from: 1  to: 3417

       1 LeuGluArgArgAspTrpSerProSer 9
         |||||||||||||||||||||||||||
    1926 CTCGAGACACGGGACTGGAGTCCCACT 1952

seq_name: /tmp1/geadir/pctseq_data/pctseq_NAIS/2_IA2_AAV[417]

seq_documentation_block:
ID   AAV80458 standard; DNA; 534720 BP.
XX
AC   AAV80458;
XX
DT   14-OCT-1998  (first entry)
XX
DE   Rhizobium species plasmid pNGR234a.
XX
KW   Symbiosis; open reading frame; ORF; plasmid vector; transposition;
KW   degradation; metabolism; host range; nitrogen fixation; nodulation;
KW   legume; plant; ds.
XX
OS   Rhizobium sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             417796..418671
FT                   /*tag= a
FT                   /standard_name  "ORF K1"
FT                   /product  "oligopeptide permease"
FT                   /note= "homologous to the oppC gene"
FT   CDS             418673..419680
FT                   /*tag= b
FT                   /standard_name  "ORF K2"
FT                   /product= "oligopeptide permease"
FT                   /note= "homologous to the oppD gene"
FT   CDS             419677..420738
FT                   /*tag= c
FT                   /standard_name  "ORF K3"
FT                   /product  "oligopeptide permease"
FT                   /note= "homologous to the oppF gene"
FT   CDS             420774..422159
FT                   /*tag= d
FT                   /standard_name  "ORF K4"
FT                   /product= "encapsulation like protein"
FT                   /note= "homologous to the CapA gene"
FT   CDS             422628..424031
FT                   /*tag= e
FT                   /standard_name  "ORF K5"
FT                   /product  "aminotransferase-like protein"
FT                   /note  "homologous to the BioA gene"
FT   CDS             424096..425594
FT                   /*tag= f
FT                   /standard_name  "ORF K6"
FT                   /product= "(gcos)aldehyde dehydrogenase-like protein"
FT   CDS             complement (426949..428028)
FT                   /*tag= g
FT                   /standard_name  "ORF K7"
FT                   /product  "transposase homologue"
FT                   /note  "homologous to the Tnp gene"
FT   CDS             428292..429623
FT                   /*tag= h
FT                   /standard_name  "ORF K8"
FT                   /product  "glutamate dehydrogenase like protein"
FT                   /note  "homologous to the GLUDI gene"
FT   CDS             complement (430548..431284)
FT                   /*tag= i
FT                   /standard_name  "ORF K9"
FT                   /product  "transposase homologue"
FT   CDS             complement (431296..432840)
FT                   /*tag= j
FT                   /standard_name= "ORF K10"
FT                   /product  "transposase homologue"
FT                   /note  "homologous to the Tnp gene"
FT   CDS             complement (433880..434110)
FT                   /*tag= k
FT                   /standard_name  "ORF K11"
FT                   /product  "protein of unknown function"
FT                   /note  "homologous to the FixU gene"
FT   CDS             complement (434107..434433)
FT                   /*tag= l
FT                   /standard_name= "ORF K12"
FT                   /product  "protein of unknown function"
FT   CDS             complement (434517..434711)
FT                   /*tag= m
FT                   /standard_name  "ORF K13"
FT                   /product  "ferredoxin/ferredoxin-like protein"
FT                   /note  "homologous to the FdxN gene"
FT   CDS             complement (434753..436234)
FT                   /*tag= n
FT                   /standard_name= "ORF K14"
FT                   /gene  "nifB"
FT                   /product  "protein involved in FeMo cofactor
FT                             biosynthesis"
FT   CDS             complement (436460..438130)
FT                   /*tag= o
FT                   /standard_name= "ORF K15"
FT                   /gene  "nifA"
FT                   /product  "positive regulator of nif, fix and other
FT                             genes"
FT   CDS             complement (438297..438590)
FT                   /*tag= p
FT                   /standard_name= "ORF K16"
FT                   /gene  "fixX"
FT                   /product= "protein required for nitrogenase activity"
FT   CDS             complement (438605..439912)
FT                   /*tag= q
FT                   /standard_name= "ORF K17"
FT                   /gene  "fixC"
FT                   /product= "protein required for nitrogenase activity"
FT   CDS             complement (439923..441042)
FT                   /*tag= r
FT                   /standard_name= "ORF K18"
FT                   /gene  "fixB"
FT                   /product= "protein required for nitrogenase activity"
FT   CDS             complement (441042..441899)
FT                   /*tag= s
FT                   /standard_name  "ORF K19"
FT                   /gene  "fixA"
```

```
FT                   /product="protein required for nitrogenase activity"
FT   CDS             complement (442316..443636)
FT                   /*tag=  t
FT                   /standard_name="ORF K2"
FT                   /product="protein of unknown function"
FT   CDS             complement (443813..443879)
FT                   /*tag=  u
FT                   /standard_name="ORF K2I"
FT                   /product="protein of unknown function"
FT   CDS             444337..445029
FT                   /*tag=  v
FT                   /standard_name="ORF K2_"
FT                   /product="Serine/Thr kinase"
FT                   /note="homologous to the NifQ gene"
FT   CDS             445088..446602
FT                   /*tag=  w
FT                   /standard_name="ORF K3"
FT                   /gene="dctA"
FT                   /product="C4 dicarboxylate transport protein"
FT                   /note="homologous to the DctA1 gene"
FT   CDS             446599..447843
FT                   /*tag=  x
FT                   /standard_name="ORF L1"
FT                   /product="cytochrome P450 like protein"
FT                   /note="homologous to the CamC gene"
FT   CDS             447844..448500
FT                   /*tag=  y
FT                   /standard_name="ORF L2"
FT                   /product="gamma-hexachlorocyclohexane dehalogenase like
FT                   protein"
FT                   /note="homologous to the LinA gene"
FT   CDS             448497..450203
FT                   /*tag=  z
FT                   /standard_name="ORF L3"
FT                   /product="putative protein with 6 predicted functions"
FT   CDS             450341..451896
FT                   /*tag=  aa
FT                   /standard_name="ORF L4"
FT                   /product="luciferase alpha subunit like protein"
FT                   /note="homologous to the LuxA gene"
FT   CDS             452980..454494
FT                   /*tag=  ab
FT                   /standard_name="ORF L6"
FT                   /gene="nifD"
FT                   /product="alpha-subunit of FeMo protein of nitrogenase"
FT   CDS             454590..456131
FT                   /*tag=  ac
FT                   /standard_name="ORF L7"
FT                   /gene="nifK"
FT                   /product="beta-subunit of FeMo protein of nitrogenase"
FT   CDS             456187..457677
FT                   /*tag=  ad
FT                   /standard_name="ORF L8"
FT                   /product="protein involved in FeMo co factor
FT                   biosynthesis"
FT                   /note="homologous to the NifB gene"
FT   CDS             457687..459096
FT                   /*tag=  ae
FT                   /standard_name="ORF L9"
FT                   /product="protein involved in FeMo co factor
FT                   biosynthesis"
FT                   /note="homologous to the FixE gene"
FT   CDS             459093..459575
FT                   /*tag=  af
FT                   /standard_name="ORF L10"
FT                   /product="protein of unknown function"
FT                   /note="homologous to the NifX gene"
FT   CDS             459579..460067
FT                   /*tag=  ag
FT                   /standard_name="ORF L11"
FT                   /product="protein of unknown function"
FT   CDS             460501..460920
FT                   /*tag=  ah
FT                   /standard_name="ORF L12"
FT                   /product="protein similar to part of the Fe protein
FT                   of nitrogenase"
FT                   /note="homologous to the NifH gene"
FT   CDS             461228..461545
FT                   /*tag=  ai
FT                   /standard_name="ORF L13"
FT                   /product="protein of unknown function"
FT   CDS             463201..464739
FT                   /*tag=  aj
FT                   /standard_name="ORF L14"
FT                   /product="peptidase like protein"
FT                   /note="homologous to the hi MPP gene"
FT   CDS             464736..466079
FT                   /*tag=  ak
FT                   /standard_name="ORF L15"
FT                   /product="processing protease like protein"
FT                   /note="homologous to the FP gene"
FT   CDS             466590..467021

alignment scores:
                Quality:     7.00              Length:       /
                 Ratio:  1.000                Gaps:         0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2 (COPY 38-46) x AAV80458/rev  ..

 Align 2:1 1/1  Score: 177404  Db: AAV80458  from: 1  to: 534720

        4 ArgArgAspArgSerProSer 9
          |||| |||||||||||||||
  177424 AGGAGGACTGCTCGTCCGTCG 177404

seq_pair  /cgi-bin/gbrs/pairseq?pairseq=pn,NA1998:DA1:AAV80459

seq_documentation block:
ID   AAV80459 standard; DNA; 536165 BP.
XX
AC   AAV80459;
XX
DT   06-JUL-1999  (first entry)
XX
DE   Rhizobium species symbiotic plasmid pNGR234.
XX
KW   Symbiosis; open reading frame; ORF; plasmid; vector; transportation;
KW   degradation; metabolism, host range, nitrogen fixation; nodulation;
KW   legume; plant; ds.
XX
OS   Rhizobium sp.
XX
PN   WO9802560-A2.
XX
PD   22-JAN-1998.
XX
PF   10-JUL-1997;   97WO-IB00950.
XX
PR   20-MAY-1997;   97GB-0010395.
PR   12-JUL-1996;   96EP-0740001.
XX
PA   (MOLE ) INST MOLECULAR BIOTECHNOLOGY.
PA   (BIOL ) LAB BIOLOGIE MOLECULAIRE PLANTES SUPERIE.
XX
PI   Broughton W.J.,  Freiberg C.,  Perret X.P.,  Rosenthal A;
XX
DR   WPI; 1998-106776/10.
XX
PT   New isolated symbiotic plasmid from Rhizobium sp. NGR234 - used to
PT   develop products for modifying plant characteristics, e.g. nitrogen
PT   fixation, synthesis of compounds and stress response
XX
PS   Claim 1; Fig 3; 228pp; English.
XX
```

```
CC   This is the nucleotide sequence of the plasmid pNGR234a isolated from
CC   Rhizobium sp. NGR234.  Open reading frames (ORF) derivable from the
CC   nucleotide sequence are claimed.  The nucleotide sequences or ORFs can
CC   be used e.g. in the transportation of compounds to and from an organism
CC   which is a host to at least one of the nucleotide sequences, ORFs or
CC   proteins, the degradation and/or metabolism of organic, inorganic,
CC   natural or xenobiotic substances in a host organism or the modification
CC   of the host range, nitrogen fixation abilities; for obtaining a synthetic
CC   minimal set of ORFs required for functional Rhizobium-legume symbiosis,
CC   especially for nodulation efficiency on host plants.
XX
SQ   Sequence 536165 BP; 111291 A; 155755 C; 157864 G; 111255 T; 0 other;


alignment_scores:
                Quality:     7.00              Length:       7
                  Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_38_46 x AAV80459/rev

 Align seq 1/1  to reverse of: AAV80459  from: 1  to: 536165

        1 ArgArgAspIleSerProSer 7
          |||||||||||||||||||||
   177424 AGGCGCGACATCGTCGCCGTCG 177404

Seq_name: /SIDS1/seqdata/genseq/genseq1_NA2000.DAT:AAV80459

Seq_documentation_block:
ID   AAV80459 standard; DNA; 536165 BP.
XX
AC   AAV80459;
XX
DT   18-MAY-2000  (first entry)
XX
DE   Rhizobium sp. NGR234 plasmid pNGR234a nucleotide sequence.
XX
KW   Rhizobium; plasmid; nucleotide sequence; ORF; open reading frame;
KW   transportation; host; organism; organic; inorganic; natural substance;
KW   xenobiotic substance; host range; nitrogen fixation; synthetic minimal
KW   set; Rhizobium-legume symbiosis; nodulation efficiency; host plant.
XX
OS   Rhizobium sp. NGR234
XX
PN   WO0027864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26055.
XX
PR   06-NOV-1998;   98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Iavitrian SV,  Teng DHF,  Simard J,  Rommens JM;
XX
DR   WPI; 2000-376481/32.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer
XX
PS   Example 5; Page 61; 157pp; English.
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
XX
```

```
SQ   Sequence 38 BP; 8 A; 12 C; 7 G; 11 T; 0 other;


alignment_scores:
                Quality:     6.00              Length:       6
                  Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_38_46 x AAA60342/rev

 Align seq 1/1  to reverse of: AAA60342  from: 1  to: 38

        1 LeuGluArgArgAspTrp 6
          |||| ||||||||||||
       24 CTGGAACGTCGTGACTGG 7

Seq_name: /SIDS1/seqdata/genseq/genseq1_NA2000.DAT:AAA60342

Seq_documentation_block:
ID   AAX11006 standard; DNA; 251 BP.
XX
AC   AAX11006;
XX
DT   30-MAR-1999  (first entry)
XX
DE   Human biallelic polymorphic DNA fragment WI-7416.
XX
KW   Polymorphism; biallelic marker; human; forensic; paternity testing; disease
KW   detection; phenotypic typing; characteristic; infection; hereditary;
KW   autoimmune disease; cancer; inflammation; drug; therapy; medicament;
KW   treatment; marker; ss.
XX
OS   Homo sapiens.
XX
PN   WO9820165-A2.
XX
PD   14-MAY-1998.
XX
PF   05-NOV-1997;   97WO-US20314.
XX
PR   06-NOV-1996;   96US-0030455.
XX
PA   (WHED ) WHITEHEAD INST BIOMEDICAL RES.
XX
PI   Hudson T,  Lander ES,  Wang D;
XX
DR   WPI; 1998-286974/25.
XX
PT   New isolated nucleic acid segments from the human genome, used for
PT   determining polymorphic forms for use in e.g. forensics, paternity
PT   testing or phenotypic typing for disease
XX
PS   Claim 1; Page 106; 310pp; English.
XX
CC   AAX10269-X12937 are human DNA fragments which contain biallelic
CC   polymorphic markers which have been isolated using the primers
CC   represented in AAX09121-X10268. The base occupying the polymorphic site
CC   is indicated by the appropriate IUPAC-IUB ambiguity code. These fragments
CC   can be used in methods for determining polymorphic forms in an individual
CC   for use in e.g. forensics, paternity testing or for phenotypic typing for
CC   diseases such as agammaglobulinemia, diabetes insipidus, Tosch-Nyhan
CC   syndrome, muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease,
CC   familial hypercholesterolemia, polycystic kidney disease, hereditary
CC   spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary
CC   haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
CC   syndrome, osteogenesis imperfecta, acute intermittent porphyria,
CC   autoimmune diseases, inflammation, cancer, diseases of the nervous
CC   system, infection by pathogenic microorganisms, and characteristics such
CC   as longevity, appearance (e.g. baldness, obesity), strength, speed,
CC   endurance, fertility, and susceptibility or receptivity to particular
CC   drugs or therapeutic treatments. The isolated polymorphic nucleic acid
CC   segments can also be used to produce medicaments for the treatment or
```

```
CC   prophylaxis of such diseases.
XX
SQ   Sequence 251 BP; 74 A; 67 C; 42 G; 67 T; 1 other;


alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_38_46 x AAX11006/rev    ..

 Align seq 1/1  to reverse of: AAX11006  from: 1  to: 251

        1 LeuGluArgArgAspTrp 6
          ||||||||||||||||||
      86 CTGGAGAGAAGGGACTGG 69

seq name: /SIDSI/seqdata/geneseq/geneseqpa/NA2000.DAT:AAC04518

seq documentation block:
ID   AAC04518 standard; cDNA; 411 BP.
XX
AC   AAC04518;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 8593.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;  99US-0122487.
XX
PA   (GEST-) GENSET.
XX
PI   Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR   WPI; 2000-500481/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures
XX
PS   Claim 1; SEQ ID 8593; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 40 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 411 BP; 88 A; 130 C; 64 G; 79 T; 1 other;
```

```
alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_38_46 x AAC04518/rev    ..

 Align seq 1/1  to reverse of: AAC04518  from: 1  to: 411

        1 LeuGluArgArgAspTrp 6
          ||||||||||||||||||
      107 CTGGAGAGAAGGGACTGG 90

seq name: /SIDSI/seqdata/geneseq/geneseqpa/NA1992.DAT:AAQ30888

seq documentation block:
ID   AAQ30888 standard; DNA; 330 BP.
XX
AC   AAQ30888;
XX
DT   02-APR-1993  (first entry)
XX
DE   Insert pMDR989-15.
XX
KW   Primer; polymerase chain reaction; amplify; 5A8; heavy chain; gp120;
KW   variable; antibody; light chain; humanised; CD4; CD4+; lymphocyte;
KW   cell surface glycoprotein; HIV induced syncytia formation; ss.
XX
OS   Synthetic.
XX
PN   WO9209305-A.
XX
PD   11-JUN-1992.
XX
PF   27-NOV-1991;  91WO-US08843.
XX
PR   27-NOV-1990;  90US-0618542.
XX
PA   (BIOJ-) BIOGEN INC.
XX
PI   Burkly LC, Chisholm PL, Rosa JJ, Rosa MD, Thomas DW;
XX
DR   WPI; 1992-198399/48.
XX
PT   New anti-CD4 antibody homologues - which bind CD4, do not block
PT   binding of HIV gp120 to CD4 but block HIV induced syncytia
PT   formation between CD4+ cells
XX
ES   Disclosure; Page 134; 205pp; English.
XX
CC   The sequences given in AAQ30888-89 are inserts which were used for
CC   humanising the 5A8 heavy chain variable region (5A8 VH). This was
CC   further accomplished by PCR mutagenesis. The humanised fragment was
CC   ligated into the vector pNN03. 5A8 VH was used in conjunction with
CC   the 5A8 light chain variable region which was also humanised using PCR
CC   mutagenesis. The antibody produced was shown to bind to CD4 but did
CC   not block the binding of gp120 to CD4. CD4 is a cell surface
CC   glycoprotein of CD4+ lymphocytes (helper/inducer cells). The antibody
CC   blocked HIV induced syncytia formation. This antibody can be used in
CC   the detection, prophylaxis and treatment of diseases caused by
CC   infective agents whose primary targets are CD4+ cells.
XX
SQ   Sequence 330 BP; 86 A; 76 C; 93 G; 75 T; 0 other;


alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_38_46 x AAQ30888    ..
```

```
 Align seq 1/1  to: AAQ30888  from: 1  to: 386

       4 ArgAspIrpSerProSer 9
         ||||||||||||||||||
     313 AGGGACTGGTCACCGTCT 330

seq_name: "SEQ Sequence Geneseq_Protein_NA" ! AC AAF67119

seq_documentation_block:
ID   AAF67119 standard; cDNA; 486 BP.
XX
AC   AAF67119;
XX
DT   09-APR-2001  (first entry)
XX
DE   Novel human polynucleotide, SEQ ID NO: 2875.
XX
KW   Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW   breast cancer; lung cancer; cancer detection; ss.
XX
OS   Homo sapiens.
XX
PN   WO200102568-A2.
XX
PD   11-JAN-2001.
XX
PF   30-JUN-2000; 2000WO-US18374.
XX
PR   02-JUL-1999;   99US-0142310.
PR   02-JUL-1999;   99US-0142311.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Williams LT,  Escobedo J,  Innis MA,  Garcia PD,  Klinger J,  Kassam A;
PI   Reinhard C,  Randazzo F,  Kennedy GC,  Fox D,  Lamson G,  Irmansah R;
PI   Orkonjakov M,  Drmanac S,  Dickson M,  Labat I,  Loshkowitz D;
PI   Kita D,  Garcia V,  Jones LW,  Strache-Crain B;
XX
DR   WPI; 2001-091805/10.
XX
PT   Library of polynucleotides for diagnosing a cancerous state of a
PT   mammalian cell and detecting cancer, particularly of the colon or
PT   prostate, comprises 3351 human polynucleotide sequences -
XX
PS   Claim 9; Page 972; 1046pp; English.
XX
CC   The present sequence is one of 3351 sequences in a library of human
CC   polynucleotides. The library is used to detect differentially expressed
CC   genes correlated with a cancerous state of a mammalian cell and can
CC   detect colon, prostate, breast and lung cancer. The library can be used
CC   to produce probes for detection of mRNA and to produce additional copies
CC   of the polynucleotides. The probes can be used for chromosome mapping of
CC   the polynucleotide and for detection of transcription levels. Ribozymes
CC   or antisense oligonucleotides can be generated. The polynucleotides and
CC   their gene products are used as genetic or biochemical markers (e.g. in
CC   blood or tissues) that will detect the earliest changes along the
CC   carcinogenesis pathway and/or monitor the efficacy of therapies and
CC   preventive interventions. The polynucleotides, polypeptides and
CC   antibodies against them can be used in pharmaceutical compositions to
CC   treat the cancers and proliferative disorders such as neoplasia,
CC   dysplasia and hyperplasia.
XX
SQ   Sequence 386 BP; 84 A; 124 C; 121 G; 57 T; 0 other;


alignment_scores:
             Quality:    6.00              Length:       6
               Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_38_46 x AAF67119/rev  ..

 Align seq 1/1  to reverse of: AAF67119  from: 1  to: 386

       3 ArgArgAspIrpSerPro 8
         |||| ||||||||||||
     347 CGTCGGGACTGGTCTCCT 330

seq_name: "SEQ Sequence Geneseq_Protein_NA" ! AC AAC79231

seq_documentation_block:
ID   AAC79231 standard; cDNA; 598 BP.
XX
AC   AAC79231;
XX
DT   05-FEB-2001  (first entry)
XX
DE   Human lung tumour-specific cDNA #184.
XX
KW   Lung tumour protein; lung cancer; cytostatic; vaccine; ss.
XX
OS   Homo sapiens.
XX
PN   WO200060077-A2.
XX
PD   12-OCT-2000.
XX
PF   30-MAR-2000; 2000WO-US08560.
XX
PR   02-APR-1999;   99US-0285823.
PR   09-AUG-1999;   99US-0370838.
PR   30-DEC-1999;   99US-0476235.
PR   03-MAR-2000; 2000US-0518809.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Reed SG,  Ludes MJ,  Mohamath R,  Seerist B;
XX
DR   WPI; 2000-638466/61.
XX
PI   Novel lung tumor polypeptides and polynucleotides, useful for
PT   detecting, monitoring or treating cancer, especially lung cancer
XX
PS   Claim 1; Page 205; 243pp; English.
XX
CC   The present sequence is given in a specification relating to compounds
CC   for therapy and diagnosis of lung cancer. Polypeptides comprising at
CC   least an immunogenic part of a lung tumour protein are disclosed.
CC   The polypeptides are useful for inhibiting the development of cancer,
CC   especially lung cancer. Samples of T cells expressing the polypeptides
CC   may be used to inhibit the development of cancer. The polypeptides are
CC   also useful for detecting and monitoring the progression of cancer,
CC   especially lung cancer.
XX
SQ   Sequence 598 BP; 140 A; 178 C; 187 G; 103 T; 0 other;


alignment_scores:
             Quality:    6.00              Length:       6
               Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_38_46 x AAC79231/rev  ..

 Align seq 1/1  to reverse of: AAC79231  from: 1  to: 598

       4 ArgAspIrpSerProSer 9
         |||| ||||||||||||
     118 CGCGACTGGAGCCCAAGT 101

seq_name: "SEQ Sequence Geneseq_Protein_NA" ! AC AAE12247
```

```
Seq documentation block:
ID   AAZ12247 standard; DNA; 711 BP.
XX
AC   AAZ12247;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria meningitidis partial ORF147 sequence.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhoea; ss.
XX
OS   Neisseria meningitidis.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999
XX
PF   09-OCT-1998;   98WO-IB001665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024198.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR ) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   P-PSDB; AAY38818.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 9; Page 455; 524pp; English.
XX
CC   Nucleotide sequences AAZ11972-Z12458 represent open reading frames
CC   (ORFs) of Neisseria meningitidis and N. gonorrhoeae which encode
CC   antigenic proteins (see AAY48499-Y38944). The antigenic proteins, their
CC   fragments, their nucleic acids and antibodies are used for diagnosis,
CC   prevention (as vaccines) or treatment of Neisseria infections,
CC   such as meningitis, septicaemia and gonorrhea. Both organisms
CC   are closely related. Fragments of the nucleic acids are useful
CC   as hybridisation probes and antisense reagents.
XX
SQ   Sequence 711 BP; 157 A; 184 C; 228 G; 142 T; 1 other;


alignment scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-copy_38_46 x AAZ12247/rc

 Align seq 1/1  to reverse of: AAZ12247  from: 1  to: 711

       4 ArgAspTrpSerProSer 9
         |||||||||||||||||||
     545 CGCGATTGGTCGCCGTCG 528

Seq name: /cgi6/ptodata/1/pda/2/us-09-126-945b-2-copy_38_46/NA2000.DAT:AAA81399

Seq documentation block
ID   AAA81399 standard; DNA; 711 BP.
XX
AC   AAA81399;
XX
```

```
DT   04-DEC-2000  (first entry)
XX
DE   N. meningitidis MenB polynucleotide sequence ORF number 75.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KW   antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW   Meningococcus B; MenB; ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO200022130-A2.
XX
PD   20-APR-2000.
XX
PF   08-OCT-1999;   99WO-US023573.
XX
PR   09-OCT-1998;   98US-0103794.
PR   30-APR-1999;   99US-0132068.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Frazer CM,  Hickey E,  Peterson J,  Tettelin H,  Venter JC;
PI   Masignani V,  Galeotti C,  Mora M,  Ratti G,  Scarselli M,  Scarlato V;
PI   Rappuoli R,  Pizza M;
XX
DR   WPI; 2000-318079/27.
XX
PT   Isolated nucleotide sequences of Neisseria meningitidis which can be
PT   used in the diagnosis and treatment of N. meningitidis infection and
PT   other Neisserial infections, for example, N.gonorrhoea
XX
PS   Disclosure; Page 218; 1760pp; English.
XX
CC   The present invention describes methods of obtaining immunogenic
CC   proteins from Neisseria genomic sequences. AAA81454 to AAA82414
CC   represent specifically claimed Neisseria meningitidis genomic DNA
CC   sequences; AAA81260 to AAA81303 and AAB25620 to AAB25664 represent
CC   Neisseria DNA sequences and their corresponding proteins; AAA81254 to
CC   AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC   isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
CC   AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC   sequences, which are all used in the exemplification of the present
CC   invention. The nucleic acid sequences, protein sequences, and antibodies
CC   against them, can be used in the manufacture of a composition. The
CC   composition can be used as a medicament (or in the manufacture of a
CC   medicament) for treating, preventing or diagnosing infection due to
CC   Neisserial bacteria. For example, some of the identified proteins could
CC   be components of vaccines against Meningococcus B; against all serotypes;
CC   and/or against all pathogenic Neissariae. Identification of sequences
CC   from the bacterium will also facilitate production of biological probes,
CC   particularly organism-specific probes. Attempts to make efficacious
CC   Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC   Multivalent vaccines have also been tried but none have successfully
CC   overcome antigenic variability. The provision of further, complete
CC   sequences may provide an opportunity to identify secreted or surface
CC   exposed proteins that may be presumed targets for the immune system and
CC   which are not antigenically variable or at least more conserved than
CC   other more variable regions.
XX
SQ   Sequence 711 BP; 157 A; 184 C; 228 G; 142 T; 1 other;


alignment scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-copy_38_46 x AAA81399/rc

 Align seq 1/1  to reverse of: AAA81399  from: 1  to: 711

       4 ArgAspTrpSerProSer 9
```

```
      ||||||||||||||||||
     545 CGCGATTGGTCGCCGTCG 528

seq_name: /SIDSl6/orf1a/genomeseq/accessdb/DNA/... AAZ12096

seq_documentation_block:
ID   AAZ12096 standard; DNA; 876 BP.
XX
AC   AAZ12096;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria meningitidis partial ORF75 sequence.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhoea; ss.
XX
OS   Neisseria meningitidis.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G;  Masignani V;  Pizza M;  Rappuoli R;  Scarlato V;
XX
DR   WPI; 1999-427407/27.
DR   P-PSDB; AAY38641.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 9; Page 196-197; 524pp; English.
XX
CC   Nucleotide sequences AAZ11972-Z12458 represent open reading frames
CC   (ORFs) of Neisseria meningitidis and N. gonorrhoeae which encode
CC   antigenic proteins (see AAY38499-Y38944). The antigenic proteins, their
CC   fragments, their nucleic acids and antibodies are used for diagnosis,
CC   prevention (as vaccines) or treatment of Neisseria infections,
CC   such as meningitis, septicaemia and gonorrhoea. Both organisms
CC   are closely related. Fragments of the nucleic acids are useful
CC   as hybridisation probes and antisense reagents.
XX
SQ   Sequence 876 BP; 190 A; 219 C; 269 G; 184 T; 14 other;


alignment_scores:
            Quality:    6.00              Length:      6
              Ratio:   1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_38_46 x AAZ12096/rev

 Align seq 1/1  to reverse of: AAZ12096  from: 1  to: 876

       4 ArgAspTrpSerProSer 9
         ||||||||||||||||||
     710 CGCGATTGGTCGCCGTCG 693

seq_name: /SIDSl6/orf1a/genomeseq/accessdb/DNA/... AAA81454
```

```
seq_documentation_block:
ID   AAA81454 standard; DNA; 876 BP.
XX
AC   AAA81454;
XX
DT   04-DEC-2000  (first entry)
XX
DE   N. meningitidis MenB polynucleotide sequence ORF number    .
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KW   antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW   Meningococcus B; MenB; ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO200022430-A2.
XX
PD   20-APR-2000.
XX
PF   08-OCT-1999;   99WO-US23574.
XX
PR   09-OCT-1998;   98US-0103794.
PR   30-APR-1999;   99US-0132068.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Frazer CM;  Hickey E;  Peterson J;  Tettelin H;  Venter JC;
PI   Masignani V;  Grandi G;  Mora M;  Ratti G;  Scarselli M;  Scarlato V;
PI   Rappuoli R;  Pizza M;
XX
DR   WPI; 2000-318079/27.
XX
PT   Isolated nucleotide sequences of Neisseria meningitidis which can be
PT   used in the diagnosis and treatment of N. meningitidis infection and
PT   other Neisserial infections, for example, N.gonorrhoea
XX
PS   Disclosure; Page 207; 1760pp; English.
XX
CC   The present invention describes methods of obtaining immunogenic
CC   proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC   represent specifically claimed Neisseria meningitidis genomic DNA
CC   sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC   Neisseria DNA sequences and their corresponding proteins; AAA81254 to
CC   AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC   isolation of Neisseria meningitidis DNA sequences, and AAA81322 to
CC   AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC   sequences, which are all used in the exemplification of the present
CC   invention. The nucleic acid sequences, protein sequences, and antibodies
CC   against them, can be used in the manufacture of a composition. The
CC   composition can be used as a medicament (or in the manufacture of a
CC   medicament) for treating, preventing or diagnosing infection due to
CC   Neisserial bacteria. For example, some of the identified proteins could
CC   be components of vaccines against Meningococcus B; against all serotypes;
CC   and/or against all pathogenic Neisseriae. Identification of sequences
CC   from the bacterium will also facilitate production of biological probes,
CC   particularly organism specific probes. Attempts to make efficacious
CC   Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC   Multivalent vaccines have also been tried but none have successfully
CC   overcome antigenic variability. The provision of further, complete
CC   sequences may provide an opportunity to identify secreted or surface
CC   exposed proteins that may be presumed targets for the immune system and
CC   which are not antigenically variable or at least more conserved than
CC   other more variable regions.
XX
SQ   Sequence 876 BP; 190 A; 219 C; 269 G; 184 T; 14 other;


alignment_scores:
            Quality:    6.00              Length:      6
              Ratio:   1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
```

US-09-126-945B-2_COPY_38_46 x AAAB1454/new ..

Align seq 1/1 to reverse of: AAAB1454 from: 1 to: 876

4 ArgAspTrpSerProSer 9
||||||||||||||||||||
710 CGCGATTGGTCGCCGTCG 693

```
PM of: US-09-126-945B-2_COPY_38_46.fas? Set 1 Result 1: NA *

Date: Sep 14, 2001  9:52 PM

About: Results were produced by the GenCore s/w/re  Version 4.5
       Copyright (c) 1993-2000 Compugen Ltd

Command line parameters:
 -MODEL frame+_p/n model -DEV xlp
 -Q "/usr2/GCG/SEQ_DIR/Queries/us_rnasat_11_s_09_126_945b_2_copy_38_46.seq"
 -DB Issued_Patents_NA -QFMT=fastap  SUFFIX olip2n.rni
 -GAPOP 4.500 -GAPEXT 0.050 -MINMATCH 0.100 -XOOPCL 0.000
 -XOOPEXT 0.000  QGAPOP 4.500 -QGAPEXT 0.050 -XGAPOP 60.000
 -XGAPEXT 60.000  FGAPOP=6.000 -FGAPEXT 7.000  YGAPOP 60.000
 -YGAPEXT 60.000  DELOP 6.000 -DELEXT=7.000  START 1
 -MATRIX=oligo -TRANS=human40.cdi -LIST 45  DOCALIGN 200
 -THR_SCORE quality  THR_MIN=1 -ALIGN 15  -MODE LOCAL -OUTFMT pir
 -NORM ext -MINLEN 0 -MAXLEN 2000000000
 -USER US=09126945B_COPY_1_61 -NCPU 6 -ICPU 1 -DISPLAY=2 -NO_XLIBXX
 -WAIT -THREADS=1

Search Information block:
Query: US-09-126-945B-2_COPY_38_46
Query length: 9
Database: Issued_Patents_NA:*
Database sequences: 824599
Database length: 94655562
Search time (sec): 322.720000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000


score_list:
Sequence            Strd Orig    ZScore   EScore Len   Documentation ..
/cgn2_6/ptodata/2/ina/5B_COMB seq US-07-916-098A-22 +     6.00  122.59  12.08      61  *
/cgn2_6/ptodata/2/ina/5B_COMB seq US-09-046-804-34      6.00  135.32  299.16    2205  *
/cgn2_6/ptodata/2/ina/5B_COMB seq US-08-933-2500-64 -     6.00  101.69  3.24 33    3520  *
/cgn2_6/ptodata/2/ina/6A_COMB seq US-09-[illegible]-34 +     6.00  [illegible]  3520  *
/cgn2_6/ptodata/2/ina/5B_COMB seq US-08-[illegible]-1 +     6.00  [illegible]  4098  *
/cgn2_6/ptodata/2/ina/[illegible]_COMB seq PCT-US96-[illegible] +     6.00  [illegible]  [illegible]
/cgn2_6/ptodata/2/ina/6A_COMB seq US-09-028-941-38 +     6.00  101.31  513.05    5006  *
/cgn2_6/ptodata/2/ina/5B_COMB seq US-08-[illegible]-2 +     6.00  [illegible]  [illegible]
/cgn2_6/ptodata/2/ina/5B_COMB seq US-08-970-269A-41      6.00  98.52  905.04    11978  *
/cgn2_6/ptodata/2/ina/5B_COMB seq US-08-970-269A-58      6.00  98.52  [illegible]  [illegible]
/cgn2_6/ptodata/2/ina/6A_COMB seq US-[illegible]-[illegible] +     6.00  [illegible]  [illegible]
/cgn2_6/ptodata/2/ina/6A_COMB seq US-08-535-057A-1 +     6.00  [illegible]  [illegible]
/cgn2_6/ptodata/2/ina/6A_COMB seq US-08-535-057A-4 +     6.00  [illegible]  [illegible]
/cgn2_6/ptodata/2/ina/6A_COMB seq US-08-535-057A-5 +     6.00  [illegible]  [illegible]
/cgn2_6/ptodata/2/ina/5B_COMB seq US-08-465-503A-10      6.00  [illegible]  [illegible]
/cgn2_6/ptodata/2/ina/5B_COMB seq US-08-465-503A-13      6.00  [illegible]  [illegible]
/cgn2_6/ptodata/2/ina/5B_COMB seq US-08-973-124-208      6.00  111.60  135.60    70  *
/cgn2_6/ptodata/2/ina/[illegible]_COMB seq PCT-US97-[illegible] +     6.00  [illegible]  [illegible]
/cgn2_6/ptodata/2/ina/6A_COMB seq US-08-535-057A-11 +     6.00  [illegible]  [illegible]
/cgn2_6/ptodata/2/ina/5B_COMB seq US-08-284-469-1 +     5.00  106.41  253.81    145  *
/cgn2_6/ptodata/2/ina/5A_COMB seq US-08-447-208A-41 +     5.00  105.11  407.27    191  *
/cgn2_6/ptodata/2/ina/6A_COMB seq US-08-184-529A-11 +     5.00  [illegible]  [illegible]
/cgn2_6/ptodata/2/ina/5A_COMB seq US-08-507-441-40 -     5.00  104.31  410.63    201  *
/cgn2_6/ptodata/2/ina/5B_COMB seq US-08-962-655A-10      6.00  101.31  410.63    201  *
/cgn2_6/ptodata/2/ina/6A_COMB seq US-09-[illegible]-[illegible] +     6.00  [illegible]  [illegible]
/cgn2_6/ptodata/2/ina/5B_COMB seq US-08-[illegible]-277-40 -     5.00  101.41  410.63    201  *
/cgn2_6/ptodata/2/ina/5B_COMB seq US-09-594-661-10      5.00  101.21  341.63    221  *
/cgn2_6/ptodata/2/ina/6B_COMB seq US-[illegible]-[illegible] +     6.00  101.28  342.07    [illegible]
/cgn2_6/ptodata/2/ina/5A_COMB seq US-07-616-470-2 -     5.00  103.92  357.83    216  *
/cgn2_6/ptodata/2/ina/5A_COMB seq US-08-341-275-2      5.00  103.92  357.89    216  *
/cgn2_6/ptodata/2/ina/6A_COMB seq US-08-[illegible]-029A-3      5.00  103.77  364.96    221  *
/cgn2_6/ptodata/2/ina/5B_COMB seq US-09-462-[illegible]-8      5.00  103.77  364.96    221  *
/cgn2_6/ptodata/2/ina/6B_COMB seq US-09-[illegible]-[illegible] +     6.00  122.79  414.26    256  *
/cgn2_6/ptodata/2/ina/6B_COMB seq US-09-218-[illegible]-14 +     5.00  102.83  439.23    271  *
/cgn2_6/ptodata/2/ina/6B_COMB seq US-[illegible]-[illegible] +     6.00  [illegible]  [illegible]  [illegible]
/cgn2_6/ptodata/2/ina/6B_COMB seq US-08-[illegible]-[illegible] +     6.00  [illegible]  [illegible]  [illegible]
/cgn2_6/ptodata/2/ina/6B_COMB seq US-08-990-824-57      5.00  100.79  525.32    345
/cgn2_6/ptodata/2/ina/6A_COMB seq US-08-535-057A-5 +     5.00  100.13  529.61    378
/cgn2_6/ptodata/2/ina/6A_COMB seq US-08-535-057A-4 +     5.00  100.11  583.57    381
/cgn2_6/ptodata/2/ina/5A_COMB seq US-08-[illegible]-029A-11 -     5.00  99.94  592.80    388
/cgn2_6/ptodata/2/ina/5B_COMB seq US-08-462-695-11      5.00  99.99  592.80    388

seq_name  /cgn2_6/ptodata/2/ina/5B_COMB seq US-07-916-098A-22

seq_documentation_block:
; Sequence 22, Application US/07916098A
; Patent No. 5871732
;  GENERAL INFORMATION:
;     APPLICANT:  BURKLY, LINDA C.
;     APPLICANT:  CHISHOLM, PATRICIA L.
;     APPLICANT:  THOMAS, DAVID W.
;     APPLICANT:  ROSA, MARGARET D.
;     APPLICANT:  ROSA, JOSEPH J.
;     TITLE OF INVENTION:  ANTI-CD4 ANTIBODY BIOMOLECULES USEFUL IN
;     TITLE OF INVENTION:  PROPHYLAXIS AND TREATMENT OF AIDS, ARC AND HIV INFECTION
;     NUMBER OF SEQUENCES:  61
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  ALLEGRETTI & WITCOFF, LTD.
;       STREET:  10 SOUTH WACKER DRIVE
;       CITY:  CHICAGO
;       STATE:  ILLINOIS
;       COUNTRY:  U.S.A.
;       ZIP:  60606
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  WORD PERFECT 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/916,098A
;       FILING DATE:  July 21, 1992
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/891/08843
;       FILING DATE:  No. 5871732ember 27, 1990
;       CLASSIFICATION:  424
;       APPLICATION NUMBER:  07/618,542
;       FILING DATE:  No. 5871732ember 27, 1990
;       CLASSIFICATION:  424
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  JOHN J. MC DONNELL
;       REGISTRATION NUMBER:  26,949
;       REFERENCE/DOCKET NUMBER:  92,410-C
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (312) 715-1000
;       TELEFAX:  (312) 715-1234
;       TELEX:  91/221-5417
;   INFORMATION FOR SEQ ID NO:  22:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  341 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  1
;       OTHER INFORMATION:  /note= "pMDR989 15 insert"
US-07-916-098A-22


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000               Gaps:        0
 Percent Similarity: 120.000   Percent Identity: 100.000
```

```
alignment block:
 US-09-126-945B-2_COPY_38_46 x US-07-916-098A-22

 Align seq 1/1  to: US-07-916-098A-22  from: 1  to: 331

        4 ArgAspTrpSerProSer 9
          |||||||||||||||||||
      314 AGGGACTGGTCAACCCTCT 331

Seq name: /cgn2_6/prod1a/2/iii_6B/pMBvacpac00/pMBvacpac00_046_894_34

Seq documentation block:
; Sequence 34, Application US/09/046,894
; Patent No. 6190857
;  GENERAL INFORMATION:
;     APPLICANT:  Ralph, David
;     APPLICANT:  An, Gang
;     APPLICANT:  O'Hara, Mark S.
;     APPLICANT:  Vetter, Robert
;     TITLE OF INVENTION:  DIAGNOSIS OF DISEASE STATE USING mRNA
;     TITLE OF INVENTION:  PROFILES IN PERIPHERAL LEUKOCYTES
;     NUMBER OF SEQUENCES:  55
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Arnold, White & Durkee
;       STREET:  P.O. Box 4433
;       CITY:  Houston
;       STATE:  Texas
;       COUNTRY:  USA
;       ZIP:  77210
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/046,894
;       FILING DATE:  Concurrently herewith
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 60/041,576
;       FILING DATE:  24 MAR 1997
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Nakashima, Richard A.
;       REGISTRATION NUMBER:  P-42,023
;       REFERENCE/DOCKET NUMBER:  UR6C.014
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (512) 418-3000
;       TELEFAX:  (512) 474-7577
;   INFORMATION FOR SEQ ID NO:  34:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4205 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-09-046-894-34


alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_38_46 x US-09-046-894-34/rev   ..

 Align seq 1/1  to reverse of: US-09-046-894-34  from: 1  to: 3205

        1 LeuGlnArgArgAspTrp 6
          ||||||||||||||||||
      802 CTGGAGCGACCGGACTGG 785

Seq name: /cgn2_6/prod1a/2/iii_6B/pMBvacpac00/pMBvacpac00_046_894_34
```

```
Seq documentation block:
; Sequence 94, Application US/08933750C
; Patent No. 5932442
;  GENERAL INFORMATION:
;     APPLICANT:  Lal, Preeti
;     APPLICANT:  Billman, Jennifer L.
;     APPLICANT:  Bandman, Olga
;     APPLICANT:  Shah, Purvi
;     APPLICANT:  Au Young, Janice
;     APPLICANT:  Yue, Henry
;     APPLICANT:  Gunther, Karl J.
;     APPLICANT:  Corley, Neil C.
;     TITLE OF INVENTION:  HUMAN REGULATORY MOLECULES
;     NUMBER OF SEQUENCES:  98
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;       STREET:  3174 Porter Drive
;       CITY:  Palo Alto
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  94304
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/933,750C
;       FILING DATE:  September 24, 1997
;       CLASSIFICATION:  536
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Billings, Lucy J.
;       REGISTRATION NUMBER:  36,749
;       REFERENCE/DOCKET NUMBER:  PF-0456 US
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415 855 0555
;       TELEFAX:  415 845-4166
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  94:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3520 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     IMMEDIATE SOURCE:
;       LIBRARY:  LIVRTUT04
;       CLONE:  2515476
US-08-933-750C-94


alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_38_46 x US-08-933-750C-94/rev   ..

 Align seq 1/1  to reverse of: US-08-933-750C-94  from: 1  to: 3520

        1 LeuGlnArgArgAspTrp 6
          ||||||||||||||||||
      405 CTGGAGCGACCGGACTGG 388

Seq name: /cgn2_6/prod1a/2/iii_6A/pMBvacpac00/pMBvacpac00_046_894_34 seq:US-09-244-613-94

Seq documentation block:
; Sequence 94, Application US/09244613
; Patent No. 6132973
```

```
;  GENERAL INFORMATION:
;     APPLICANT:  Lal, Preeti
;     APPLICANT:  Hillman, Jennifer L.
;     APPLICANT:  Bandman, Olga
;     APPLICANT:  Shah, Purvi
;     APPLICANT:  Au Young, Janice
;     APPLICANT:  Yue, Henry
;     APPLICANT:  Gueqler, Karl J.
;     APPLICANT:  Corley, Neil C.
;     TITLE OF INVENTION:  HUMAN REGULATORY MOLECULES
;     NUMBER OF SEQUENCES:  98
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;       STREET:  3174 Porter Drive
;       CITY:  Palo Alto
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  94304
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/234,613
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/933,750
;       FILING DATE:  September 23, 1997
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Billings, Lucy J.
;       REGISTRATION NUMBER:  36,749
;       REFERENCE/DOCKET NUMBER:  PF-0356 US
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415-855-0555
;       TELEFAX:  415-845-4166
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  94:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4520 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     IMMEDIATE SOURCE:
;       LIBRARY:  LIVRTUT04
;       CLONE:  2515476
US-09-234-613-94


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_38_46 x US-09-234-613-94;rev    .

 Align seq 1/1  to reverse of: US-09-234-613-94  from: 1  to: 4520

        1 LeuGluArgArgAspTrp 6
          ||||||||||||||||||
     405 CTGGAGCGACGGGACTGG 388

seq_name: /cgn2_6/ptodata/2/iia/58_?.MB ; pj ?? ?? 417 410A 1

seq_documentation_block:
; Sequence 1, Application US/08317410A
; Patent No. 5858701
;  GENERAL INFORMATION:
;     APPLICANT:  WHITE, Morris F.
;     APPLICANT:  SUN, Xiao Jian
;     APPLICANT:  PIERCE, Jacalyn B.
;     TITLE OF INVENTION:  THE IRS FAMILY OF GENES
;     NUMBER OF SEQUENCES:  64
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  LAHIVE & COCKFIELD
;       STREET:  28 State Street
;       CITY:  Boston
;       STATE:  Massachusetts
;       COUNTRY:  USA
;       ZIP:  02109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  ASCII text
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/317,410A
;       FILING DATE:  03-OCT-1994
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Louis Myers
;       REGISTRATION NUMBER:  35,965
;       REFERENCE/DOCKET NUMBER:  IRP 022
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617)227-7400
;       TELEFAX:  (617)227 5941
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4088 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  60..4022
US-08-317-410A 1


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_38_46 x US-08-317-410A-1    .

 Align seq 1/1  to: US-08-317-410A-1  from: 1  to: 4088

        3 ArgArgAspTrpSerPro 8
          ||| ||||||||||||
     2539 AGGAGACGGGACTGGAGCCC 2556

seq_name: /cgn2_6/ptodata/2/iia/PCTUS_COMB.seq:PCT-US95-13041-1

seq_documentation_block:
; Sequence 1, Application PCTUS9513041
;  GENERAL INFORMATION:
;     APPLICANT:  WHITE, Morris F.
;     APPLICANT:  SUN, Xiao Jian
;     APPLICANT:  PIERCE, Jacalyn B.
;     TITLE OF INVENTION:  THE IRS FAMILY OF GENES
;     NUMBER OF SEQUENCES:  64
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  LAHIVE & COCKFIELD
;       STREET:  60 State Street, Suite 510
;       CITY:  Boston
;       STATE:  Massachusetts
;       COUNTRY:  USA
;       ZIP:  02109 1875
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;       SOFTWARE:  ASCII text
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/13041
;       FILING DATE:  Herewith
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/317,410
;       FILING DATE:  04-OCT-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Louis Myers
;       REGISTRATION NUMBER:  35,965
;       REFERENCE/DOCKET NUMBER:  JBH 0225
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617)227-7400
;       TELEFAX:  (617)227-5941
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4088 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  69..4022
PCT US95 13041 1


alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_38_46 x PCT US95 13041 1    ..

 Align seq 1/1   to: PCT US95 13041 1  from: 1  to: 4088

        4 ArgArgAspTrpSerPro 8
          ||||||||||||||||||
     2539 AGGAGAGACTGGAGCCCC 2556

Seq name: /cgn2_6/ptodata/2/ina/6A_COMB seq:US-09-028-934-33

Seq documentation block:
; Sequence 33, Application US/09028934
; Patent No. 6117670
;  GENERAL INFORMATION:
;     APPLICANT:  Ligon, James M.
;     APPLICANT:  Hill, Dwight S.
;     APPLICANT:  Lam, Steven T.
;     APPLICANT:  Hammer, Philip E.
;     APPLICANT:  Van Dyk, Karl Hell
;     APPLICANT:  Kirner, Sabine
;     APPLICANT:  Young, Thomas R.
;     TITLE OF INVENTION:  Pyrrolnitrin Biosynthesis Genes and Uses
;     TITLE OF INVENTION:  Thereof
;     NUMBER OF SEQUENCES:  47
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  No. 6117670artis Corporation
;       STREET:  3054 Cornwallis Road
;       CITY:  Research Triangle Park
;       STATE:  NC
;       COUNTRY:  USA
;       ZIP:  27709
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/028,934
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/729,214
;       FILING DATE:  09-OCT-1996
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/258,261
;       FILING DATE:  08-JUN-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Meiss, J. Timothy
;       REGISTRATION NUMBER:  38,241
;       REFERENCE/DOCKET NUMBER:  [illegible]
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  919 541 8587
;       TELEFAX:  919 541 8689
;   INFORMATION FOR SEQ ID NO:  33:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  5996 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Myxococcus fulvus
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  174..1284
;       OTHER INFORMATION:  /product= "PrnB"
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1341..3038
;       OTHER INFORMATION:  /product= "PrnC"
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  3100..4239
;       OTHER INFORMATION:  /product= "PrnD"
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  complement (4258..5880)
;       OTHER INFORMATION:  /product= "PrnA"
US-09-028-934-33


alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_38_46 x US-09-028-934-33    ..

 Align seq 1/1   to: US-09-028-934-33  from: 1  to: 5996

        4 ArgAspTrpSerProSer 9
          |||||||||||||||||
     437 AGGGACTGGAGCCCATCC 454

Seq name: /cgn2_6/ptodata/2/ina/6B_COMB seq:US-08-970-269A-32

Seq documentation block:
; Sequence 32, Application US/08970269A
; Patent No. 5976803
;  GENERAL INFORMATION:
;     APPLICANT:  Kathryn Meek
;     TITLE OF INVENTION:  Genetic Test For Equine Severe
;     TITLE OF INVENTION:  Combined Immunodeficiency Disease
;     NUMBER OF SEQUENCES:  32
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Dr. Benjamin A. Adler
;       STREET:  8011 Candle Lane
;       CITY:  Houston
;       STATE:  Texas
;       COUNTRY:  USA
```

```
;       ZIP:  77071
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  Apple
;       OPERATING SYSTEM:  Macintosh
;       SOFTWARE:  Microsoft Word for Macintosh
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/970,269A
;       FILING DATE:  No. 5976803ember 14, 1997
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Adler Ph.D., Benjamin A.
;       REGISTRATION NUMBER:  35,423
;       REFERENCE/DOCKET NUMBER:  D5860
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  713-777-2321
;       TELEFAX:  713-777-6908
;   INFORMATION FOR SEQ ID NO:  32:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  11873 bp
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double stranded
;       TOPOLOGY:  linear
;     MOLECULE TYPE:
;       DESCRIPTION:  other nucleic acid
;     HYPOTHETICAL:  no
;     ANTI-SENSE:  no
;     ORIGINAL SOURCE:
;     FEATURE:
US-08-970-269A-32


alignment_scores:
             Quality:     6.00              Length:       6
               Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_38_46 x US-08-970-269A-32/rc   .

 Align seq 1/1  to reverse of: US-08-970-269A-32  from: 1  to: 11873

        4 ArgAspTrpSerProSer 9
          ||||||||||||||||||
     8677 AGGGACTGGTCTCCCTCA 8660

seq_name:  [illegible]

seq_documentation_block:
; Sequence 32, Application US/08970269A
; Patent No. 5976803
;  GENERAL INFORMATION:
;     APPLICANT:  Katheryn Meek
;     TITLE OF INVENTION:  Genetic Test For Equine Severe
;     TITLE OF INVENTION:  Combined Immunodeficiency Disease
;     NUMBER OF SEQUENCES:  32
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Dr. Benjamin A. Adler
;       STREET:  8011 Candle Lane
;       CITY:  Houston
;       STATE:  Texas
;       COUNTRY:  USA
;       ZIP:  77071
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  Apple
;       OPERATING SYSTEM:  Macintosh
;       SOFTWARE:  Microsoft Word for Macintosh
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/970,269A
;       FILING DATE:  No. 5976803ember 14, 1997
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Adler Ph.D., Benjamin A.
;       REGISTRATION NUMBER:  35,423
;       REFERENCE/DOCKET NUMBER:  D5860
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  713-777-2321
;       TELEFAX:  713-777-6908
;   INFORMATION FOR SEQ ID NO:  31:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  11878 bp
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double stranded
;       TOPOLOGY:  linear
;     MOLECULE TYPE:
;       DESCRIPTION:  other nucleic acid
;     HYPOTHETICAL:  no
;     ANTI-SENSE:  no
;     ORIGINAL SOURCE:
;     FEATURE:
US-08-970-269A-31


alignment_scores:
             Quality:     6.00              Length:       6
               Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_38_46 x US-08-970-269A-31/rc   .

 Align seq 1/1  to reverse of: US-08-970-269A-31  from: 1  to: 11878

        4 ArgAspTrpSerProSer 9
          ||||||||||||||||||
     8677 AGGGACTGGTCTCCCTCA 8660

seq_documentation_block:
; Sequence 31, Application US/08970269A
; Patent No. 5976803
;  GENERAL INFORMATION:
;     APPLICANT:  Katheryn Meek
;     TITLE OF INVENTION:  Genetic Test For Equine Severe
;     TITLE OF INVENTION:  Combined Immunodeficiency Disease
;     NUMBER OF SEQUENCES:  32
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Dr. Benjamin A. Adler
;       STREET:  8011 Candle Lane
;       CITY:  Houston
;       STATE:  Texas
;       COUNTRY:  USA
;       ZIP:  77071
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  Apple
;       OPERATING SYSTEM:  Macintosh
;       SOFTWARE:  Microsoft Word for Macintosh
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/970,269A
;       FILING DATE:  No. 5976803ember 14, 1997
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Adler Ph.D., Benjamin A.
;       REGISTRATION NUMBER:  35,423
;       REFERENCE/DOCKET NUMBER:  D5860
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  713-777-2321
;       TELEFAX:  713-777-6908
;   INFORMATION FOR SEQ ID NO:  28:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  11883 bp
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double stranded
```

```
;       TOPOLOGY:  linear
;     MOLECULE TYPE:
;       DESCRIPTION:  other nucleic acid
;     HYPOTHETICAL:  no
;     ANTI-SENSE:  no
;     ORIGINAL SOURCE:
US 08/970,269A 28


alignment scores:
              Quality:     6.00              Length:       6
                Ratio:  1.000                Gaps:         0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_38_46 x US 08/970,269A 28/res    ..

 Align seq 1/1  to reverse of: US 08/970,269A 28  from: 1  to: 11884

        4 ArgAspTrpSerProSer 9
          ||||||||||||||||||
     8681 AGGGACTGGTCTCCCTCA 8664

Seq_name: /cpa2_ptsdata/seq_loc_5A_COMb.seq.... 08_59_281_5

seq documentation block:
; Sequence 6, Application US/08159784
; Patent No. 5644784
;  GENERAL INFORMATION:
;     APPLICANT:  Bjorn R. Olsen
;     TITLE OF INVENTION:  NOVEL COLLAGEN AND USES THEREOF
;     NUMBER OF SEQUENCES:  9
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Fish & Richardson
;       STREET:  225 Franklin Street
;       CITY:  Boston
;       STATE:  Massachusetts
;       COUNTRY:  U.S.A.
;       ZIP:  02110-2804
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  4.5" Diskette, 1.44 Mb
;       COMPUTER:  IBM PS/2 Model 50Z or 55SX
;       OPERATING SYSTEM:  MS-DOS (Version 5.0)
;       SOFTWARE:  WordPerfect (Version 5.1)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/159,784
;       FILING DATE:  December 1, 1993
;       CLASSIFICATION:  530
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  John F. Freeman
;       REGISTRATION NUMBER:  29,066
;       REFERENCE/DOCKET NUMBER:  00246/170001
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 542 5070
;       TELEFAX:  (617) 542 8906
;       TELEX:  200154
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  18
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US 08/159,784 6


alignment scores:
              Quality:     5.00              Length:       5
                Ratio:  1.000                Gaps:         0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_38_46 x US 08/159,784 6/res    ..

 Align seq 1/1  to reverse of: US 08/159,784 6  from: 1  to: 18

        3 ArgArgAspTrpSer 7
          ||| |||||||||||
       17 CGCAGGGACTGGAGC 3

seq name: /cpa2_ptsdata/seq_loc_5A_COMb.seq.US 08_595_057A_10

seq documentation block:
; Sequence 10, Application US/08595057A
; Patent No. 6140104
;  GENERAL INFORMATION:
;     APPLICANT:  LEROY-DUS, DIDIER
;     APPLICANT:  AGAISSE, HERVE
;     TITLE OF INVENTION:  NUCLEOTIDE SEQUENCES FOR THE CONTROL OF
;     TITLE OF INVENTION:  THE EXPRESSION OF DNA SEQUENCES IN A CELL HOST
;     NUMBER OF SEQUENCES:  17
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
;       ADDRESSEE:  P.C.
;       STREET:  1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
;       CITY:  ARLINGTON
;       STATE:  VA
;       COUNTRY:  USA
;       ZIP:  22202
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/595,057A
;       FILING DATE:  20 DEC 1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  FR/FR91/00525
;       FILING DATE:  05-MAY 1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  FR 93/05387
;       FILING DATE:  05 MAY 1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  OBLON, NORMAN F.
;       REGISTRATION NUMBER:  24,618
;       REFERENCE/DOCKET NUMBER:  660-103-0X PCT
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  703 413 3000
;       TELEFAX:  703 413 2220
;   INFORMATION FOR SEQ ID NO:  10:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  32 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US 08/595,057A 10


alignment scores:
              Quality:     5.00              Length:       5
                Ratio:  1.000                Gaps:         0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_38_46 x US 08/595,057A 10    ..

 Align seq 1/1  to: US 08/595,057A 10  from: 1  to: 32

        1 LeuGluArgArgAsp 5
          |||||||||||||||
        1 CTGGAAAGGAGGGAT 15
```

```
seq_name: /cgn2_6/ptodata/2/iea/2_COPY_38_46_OMB_seq_US-08-545-057A-9

seq_documentation_block:
; Sequence 9, Application US/08545057A
; Patent No. 6140104
; GENERAL INFORMATION:
;   APPLICANT: LERECLUS, DIDIER
;   APPLICANT: AGAISSE, HERVE
;   TITLE OF INVENTION: NUCLEOTIDE SEQUENCES FOR THE CONTROL OF
;   TITLE OF INVENTION: THE EXPRESSION OF DNA SEQUENCES IN A CELL HOST
;   NUMBER OF SEQUENCES: 17
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
;     ADDRESSEE: P.C.
;     STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
;     CITY: ARLINGTON
;     STATE: VA
;     COUNTRY: USA
;     ZIP: 22202
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/545,057A
;     FILING DATE: 20-DEC-1995
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: PCT/FR94/00525
;     FILING DATE: 05-MAY-1994
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: FR 93/05387
;     FILING DATE: 05-MAY-1993
;   ATTORNEY/AGENT INFORMATION:
;     NAME: OBLON, NORMAN F.
;     REGISTRATION NUMBER: 24,618
;     REFERENCE/DOCKET NUMBER: 660-103-0X PCT
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 703-413-3000
;     TELEFAX: 703-413-2220
;  INFORMATION FOR SEQ ID NO: 9:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 38 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: double
;     TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
US-08-545-057A-9


alignment_scores:
            Quality:     5.00              Length:       5
              Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_38_46 x US-08-545-057A-9    ..

 Align seq 1/1  to: US-08-545-057A-9  from: 1  to: 38

        1 LeuGluArgArgAsp 5
          |||||||||:|||||
        4 CTGGAAAGGAGGGAT 17

seq_name: /cgn2_6/ptodata/2/iea/2_COPY_38_46_OMB_seq_US-09-262-773-76

seq_documentation_block:
; Sequence 76, Application US/09262773
; Patent No. 6225451
; GENERAL INFORMATION:
;   APPLICANT: Ballinger, Dennis G.
;   APPLICANT: Ding, Wei
;   APPLICANT: Wagner, Susanne
;   APPLICANT: Boss, Mark A.
;   TITLE OF INVENTION: CHROMOSOME 11-LINKED CORONARY HEART DISEASE
;   TITLE OF INVENTION: SUSCEPTIBILITY GENE CHD1
;   FILE REFERENCE: Myriad 3
;   CURRENT APPLICATION NUMBER: US/09/262,773
;   CURRENT FILING DATE: 1999-03-04
;   NUMBER OF SEQ ID NOS: 210
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 76
;   LENGTH: 38
;   TYPE: DNA
;   ORGANISM: primer
US-09-262-773-76


alignment_scores:
            Quality:     5.00              Length:       5
              Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_38_46 x US-09-262-773-76/rev    ..

 Align seq 1/1  to reverse of: US-09-262-773-76  from: 1  to: 38

        2 GluArgArgAspTrp 6
          |||||:|||||||||
       21 GAGCGTCGTGACTGG 7

seq_name: /cgn2_6/ptodata/2/iea/2_COPY_38_46_OMB_seq_US-09-262-773-185

seq_documentation_block:
; Sequence 185, Application US/09262773
; Patent No. 6225451
; GENERAL INFORMATION:
;   APPLICANT: Ballinger, Dennis G.
;   APPLICANT: Ding, Wei
;   APPLICANT: Wagner, Susanne
;   APPLICANT: Boss, Mark A.
;   TITLE OF INVENTION: CHROMOSOME 11-LINKED CORONARY HEART DISEASE
;   TITLE OF INVENTION: SUSCEPTIBILITY GENE CHD1
;   FILE REFERENCE: Myriad 3
;   CURRENT APPLICATION NUMBER: US/09/262,773
;   CURRENT FILING DATE: 1999-03-04
;   NUMBER OF SEQ ID NOS: 210
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 185
;   LENGTH: 38
;   TYPE: DNA
;   ORGANISM: primer
US-09-262-773-185


alignment_scores:
            Quality:     5.00              Length:       5
              Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_38_46 x US-09-262-773-185/rev    ..

 Align seq 1/1  to reverse of: US-09-262-773-185  from: 1  to: 38

        2 GluArgArgAspTrp 6
          |||||:|||||||||
       21 GAGCGTCGTGACTGG 7
```

```
OM of: US-09-126-945B-2_COPY_38_46 to: EST:*   ent_format : pir

Date: Sep 14, 2001  7:47 PM

About: Results were produced by the GenCore software, version 4.5.
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=framet_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO-spool/US09126945/runat_11.rg.40.1.511_1/app_query.fasta_1_1967
 -DB EST -QFMT=fastap -SUFFIX=olip2n.rst -GAP=8 4.500
 -GAPEXT 0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT 0.000
 -QGAPOP 4.500 -qGAPEXT=0.050 -xGAPOP=60.000 -XGAPEXT 60.000
 -PGAPOP 6.000 -PGAPEXT=7.000  YGAPOP=60.000 -YGAPEXT 60.000
 -DELOP 6.000 -DELEXT=7.000 -START=1 -MATRIX oligo
 -TRANS human40.cdi -LIST 45  DOCALIGN 200  THR_SCORE quality
 -THR_MIN 1  ALIGN 15  MODE LOCAL -OUTFMT pir -NORM ext  MINLEN 0
 -MAXLEN 2000000000 -USER=US09126945_@CGN1_1_4645 -NCPU=6
 -TOPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS 1

Search Information Block:
Query: US-09-126-945B-2_COPY_38_46
Query length: 9
Database: EST:*
Database sequences: 10228115
Database length: 4311459454
Search time (sec): 9087.630000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000


score_list:
Sequence              Strd orig     ZScore    EScore Len  ! Documentation ...
qb_est24:AI685592       +     9.00  175.02      1.2e-07  386  ! AI685592 tg20c07.x1 NCI_CGAP_Pr
qb_est69:BG428411       +     9.00  169.64      2.4e-07  966  ! BG428411 602341771F1 NIH_MGC_37
qb_est76:BE563288       +     9.00  169.41      2.5e-07  923  ! BE563288 601345418F1 NIH_MGC_49
qb_est80:BE871340       +     9.00  169.33      2.5e-07  945  ! BE871340 601449457F1 NIH_MGC_65
qb_est97:BG179568       +     9.00  169.02      2.6e-07  991  ! BG179568 602328182F1 NIH_MGC_91
qb_qss20:AZ018772       +     7.00  138.31      138.43   607  ! AZ018772 RPCI 23_334M15.TV RPCI
qb_est9:AA587454        +     7.00  137.85      156.56   457  ! AA587454 nn84a10.s1 NCI_CGAP_Br
qb_qss2?:AZ018678       +     7.00  137.14      162.85   669  ! AZ018678 RPCI 23_331K19.TV RPCI
qb_est15:AI028205       +     7.00  136.45      176.61   333  ! AI028205 ow61e01.r1 Soares_fetal
qb_est19:AI369207       +     7.00  135.83      198.15   454  ! AI369207 qy75d06.x1 NCI_CGAP_Br
qb_est93:BF879453       +     7.00  135.41      209.79   184  ! BF879453 QV3-BT0171-241100-494
qb_est45:AW317978       +     7.00  135.21      215.62   583  ! AW317978 xj96f07.x1 NCI_CGAP_Ut
qb_est70:BE105714       +     7.00  134.97      212.47   519  ! BE105714 QV1-R-B01-asi-f-09-0-UI
qb_est89:BF563188       +     7.00  134.89      228.70   568  ! BF563188 UI-R-B01-ajn-c-05-0-UI
qb_est85:BF895418       +     7.00  134.41      241.34   576  ! BF895418 EST450899 Rat Osteo Ho
qb_qss27:AZ464747       +     7.00  134.41      241.34   576  ! AZ464747 IMAGE3274134 Mouse 10Kb
qb_qss3:CN502701        +     7.00  134.24      244.29   582  ! AL185139 Tetraodon nigroviridis
qb_qss10:AQ362110       -     7.00  133.13      269.01   693  ! AQ362110 mqxb0004F10r COGI Rice
qb_est40:AV708973       -     7.00  133.07      271.22   703  ! AV708973 AV708973 ADC Homo sapi
qb_qss4:CN504HWS        +     7.00  131.65      325.39   875  ! AL204503 Tetraodon nigroviridis
qb_est96:BG110172       -     7.00  131.55      329.33   888  ! BG110172 602227810F1 NIH_MGC_80
qb_est99:BG343061       -     7.00  131.49      332.05   897  ! BG343061 HVSMEq0001e001f Hordeum
qb_qss5:CN505MY9        +     7.00  129.77      413.89  1175  ! AL344682 Tetraodon nigroviridi
qb_est4:AA223893        -     6.00  123.74      897.08   151  ! AA223893 zr10q08.r1 Stratagene
qb_est85:BF222161       -     6.00  123.14      969.16   166  ! BF222161 7p44d06.x1 NCI_CGAP_Pr
qb_qss31:AZ717561       +     6.00  123.02      983.43   169  ! AZ717561 RPCI 24_72b24.TV RPCI
qb_est46:AW381617       -     6.00  122.87      1.0e+03  174  ! AW381617 QV0-HT0309-071299-067
qb_est14:AA314423       +     6.00  122.59      1.0e+03  181  ! AA314423 zz81c01.r1 Soares_mamm
qb_est4:AA271352        +     6.00  122.41      1.1e+03  186  ! AA271352 vb72a06.r1 Soares_mous
qb_est50:AW699528       +     6.00  122.41      1.1e+03  186  ! AW699528 qb06q10.y1 Moss EST li
qb_est100:BG405780      -     6.00  122.21      1.1e+03  192  ! BG405780 sac37b04.y1 Gm c1051 (
qb_est8:AA510860        +     6.00  121.64      1.2e+03  210  ! AA510860 vh55c09.r1 Soares_mamm
qb_qss19:AQ384782       +     6.00  121.52      1.2e+03  214  ! AQ384782 RPC111 148N13.TV RPC1
qb_est46:AW452392       +     6.00  121.49      1.2e+03  215  ! AW452392 CM1-BT0141 172000-013
qb_est31:AV010798       +     6.00  121.09      1.3e+03  229  ! AV010798 AV010798 Mus musculus
qb_est68:BB594582       -     6.00  121.09      1.3e+03  229  ! BB594582 BB594582 RIKEN full le
qb_est22:AI563489       +     6.00  120.92      1.3e+03  235  ! AI563489 EST00613 watermelon le
qb_est10:C39596         +     6.00  120.79      1.3e+03  240  ! C39596 C39596 Yuji Kohara unpub
qb_est68:BB600823       +     6.00  120.74      1.3e+03  242  ! BB600823 BB600823 RIKEN full le
qb_est72:BE247789       -     6.00  120.66      1.3e+03  245  ! BE247789 EAAE2B423 Endosperm
qb_est90:BF606763       -     6.00  120.63      1.3e+03  246  ! BF606763 273898 MARC 60V Ho
qb_qss6:AQ286790        +     6.00  120.58      1.3e+03  248  ! AQ286790 RPCI11 23L20.TV RPC
qb_est43:AW127000       -     6.00  120.56      1.3e+03  249  ! AW127000 qa1eq01.y1 Moss EST
qb_est86:BF874927       -     6.00  120.41      1.4e+03  255  ! BF874927 BNS-SD171-071099-0
qb_est93:BF853443       -     6.00  120.28      1.4e+03  260  ! BF853443 MR3-EN0090-201200-0

seq_name: qb_est24:AI685592

seq_document_main_block:
LOCUS       AI685592          386 bp    mRNA            EST       15-FEB-1999
DEFINITION  tg20c07.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2251596 3'
            similar to contains PIR5:I3 MSR1 repetitive element ;, mRNA
            sequence.
ACCESSION   AI685592
VERSION     AI685592.1  GI:4896886
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 386)
  AUTHORS   NCI_CGAP_Pr28, Soares,M.B., Bonaldo,M.F., Jin,L. and Lennon,G.G.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael B.
            Ernest S.Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov/image/html/
            Insert Length: 1143  Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 480.
FEATURES             Location/Qualifiers
     source          1..386
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2251596"
                     /clone_lib="NCI_CGAP_Pr28"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3-Pac (Pharmacia)
                     with a modified polylinker; Plasmid DNA from the
                     normalized library NCI_CGAP_Pr22 was prepared, and ss
                     circles were made in vitro. Following PAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR amplified cDNAs from a pool
                     of 5,000 clones made from the same library (clonelDs
                     985608-986759, 1101192-1101959, and 1217028-1220615).
                     Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT        72 a    146 c    110 g     64 t
ORIGIN


alignment_scores:
            Quality:     9.00              Length:       9
              Ratio:    1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 QS 39 tgcgtgatg 47  qb_est24:AI685592 x AI685592  ..

 Align s: 1/1  to: AI685592  from: 1  to: 386

       1 LeuGluArgArgAspIlpSerProSer 9
         |||||||||||||||||||||||||||
```

```
     272 CTCGAGAGACGGGACTGGAGTCCCAGT 298

seq_name: qb_est99:BG328411

seq_documentation_block:
LOCUS       BG328411        899 bp    mRNA            EST       27-SEP-2001
DEFINITION  602427749F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4546860 5',
            mRNA sequence.
ACCESSION   BG328411
VERSION     BG328411.1  GI:13134849
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 899)
  AUTHORS   NIH_MGC http://mgc.nih.gov.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: NIH Intramural Sequencing Center
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1284  row: b  column: 13
            High quality sequence stop: 602.
FEATURES             Location/Qualifiers
     source          1..899
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4546860"
                     /clone_lib="NIH_MGC_15"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage resistant)"
                     /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >900bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT      163 a    346 c    277 g    113 t
ORIGIN


alignment scores:
            Quality:     9.00              Length:       9
              Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2 COPY_38_46 x BG328411    ..

 Align seq 1/1  to: BG328411  from: 1  to: 899

        1 LeuGluArgArgAspTrpSerProSer 9
          |||||||||||||||||||||||||||
      524 CTCGAGAGACGGGACTGGAGTCCCAGT 550

seq_name: qb_est76:BE563288

seq_documentation_block:
LOCUS       BE563288        933 bp    mRNA            EST       15-AUG-2000
DEFINITION  601335410F1 NIH_MGC_49 Homo sapiens cDNA clone IMAGE:3689402 5',
            mRNA sequence.
ACCESSION   BE563288
VERSION     BE563288.1  GI:9807100
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 933)
  AUTHORS   NIH_MGC http://mgc.nih.gov.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM384  row: c  column: 03
            High quality sequence stop: 678.
FEATURES             Location/Qualifiers
     source          1..933
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3689402"
                     /clone_lib="NIH_MGC_49"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B (phage resistant)"
                     /note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G).  Library constructed
                     by Ling Hong in the laboratory of Gerald M. Rubin
                     (University of California, Berkeley) using ZAP cDNA
                     synthesis kit (Stratagene) and Superscript II RT (Life
                     Technologies)."
BASE COUNT      148 a    316 c    319 g    150 t
ORIGIN


alignment scores:
            Quality:     9.00              Length:       9
              Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2 COPY_38_46 x BE563288    ..

 Align seq 1/1  to: BE563288  from: 1  to: 933

        1 LeuGluArgArgAspTrpSerProSer 9
          |||||||||||||||||||||||||||
      507 CTCGAGAGACGGGACTGGAGTCCCAGT 533

seq_name: qb_est80:BE871340

seq_documentation_block:
LOCUS       BE871340        945 bp    mRNA            EST       20-OCT-2000
DEFINITION  601449457F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3853344 5',
            mRNA sequence.
ACCESSION   BE871340
VERSION     BE871340.1  GI:10320116
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 945)
  AUTHORS   NIH_MGC http://mgc.nih.gov.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
```

```
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9577  row: a  column: 15
            High quality sequence stop: 658.
FEATURES             Location/Qualifiers
     source          1..945
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4853334"
                     /clone_lib="NIH_MGC_65"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: oligo dT.
                     Average insert size 1.8 Kb.  Library constructed by Life
                     Technologies.  "
BASE COUNT       170 a     364 c     300 g     112 t
ORIGIN


alignment_scores:
            Quality:    9.00             Length:       9
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_38_46 x BE871340    .

 Align seq 1/1  to: BE871340  from: 1  to: 945

       1 LeuGluArgArgAspTrpSerProSer 9
         ||||||||||||||||||||||||||||
     474 CTCGAGAGACGGGACTGGAGTCCCAGT 500

seq_name: gb_est97:BE871340

seq_documentation_block:
LOCUS        BE871340       991 bp    mRNA            EST       06-FEB-2001
DEFINITION  602329182F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4429413 5',
            mRNA sequence.
ACCESSION   BE871340
VERSION     BE871340.1  GI:12686271
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo
REFERENCE   1  (bases 1 to 991)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTB
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10181  row: d  column: 22
            High quality sequence stop: 655.
FEATURES             Location/Qualifiers
     source          1..991
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4429413"
                     /clone_lib="NIH_MGC_91"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: prostate; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Oligo dT primed.
                     Average insert size 1.4 Kb.  Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT       190 a     351 c     305 g     145 t
ORIGIN


alignment_scores:
            Quality:    9.00             Length:       9
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_38_46 x BG179568    .

 Align seq 1/1  to: BG179568  from: 1  to: 991

       1 LeuGluArgArgAspTrpSerProSer 9
         ||||||||||||||||||||||||||||
     15 CTCGAGAGACGGGACTGGAGTCCCAGT 41

seq_name: gb_gss20:AZ018772

seq_documentation_block:
LOCUS        AZ018772       307 bp    DNA             GSS       25-FEB-2000
DEFINITION  RPCI-23-334M19.TV RPCI-23 Mus musculus genomic clone RPCI-23-334M19
            , DNA sequence.
ACCESSION   AZ018772
VERSION     AZ018772.1  GI:7094156
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 307)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Lovins,M., Mcgann,S., Isogaye,G., Geer,K., Krol,M., de Jong,P
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-23-334M19.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23, for BAC
            library availability, please contact Pieter de Jong
            (pieter.dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderinfoframe.htm)
            or from Research Genetics (info@resgen.com). BAC end pairs:
            http://www.tigr.org/tdb/mgc/mouse_bac_end_intro.html
            Plate: 334  row: M  column: 19
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..307
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-334M19"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: kidney, Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and
                     brain genomic DNA was isolated and partially digested
                     with a combination of EcoRI and EcoRI Methylase.  Size
```

selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies). "
BASE COUNT 92 a 84 c 64 g 68 t
ORIGIN

alignment scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2_COPY_38_46 x AZ018772 ..

Align seq 1/1 to: AZ018772 from: 1 to: 397

3 ArgArgAspTrpSerProSer 9
||||||||||||||||||||
271 AGGAGAGACTGGAGTCCTTCA 291

seq_name: qb_est9:AA587454

seq_documentation_block:
LOCUS AA587454 457 bp mRNA EST 12-SEP-1997
DEFINITION nn84a10.s1 NCI_CGAP_Br2 Homo sapiens cDNA clone IMAGE:1097850 3',
mRNA sequence.
ACCESSION AA587454
VERSION AA587454.1 GI:2398268
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 457)
AUTHORS NCI_CGAP http://www.ncbi.nlm.nih.gov/ncicgap
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Seq primer: 40ml3 fwd, ET from Amersham
High quality sequence stop: 414.
FEATURES Location/Qualifiers
source 1..457
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1097850"
/clone_lib="NCI_CGAP_Br2"
/sex="female, pooled"
/tissue_type="breast"
/lab_host="DH10B"
/note="Vector: pT7T3-Pac (Pharmacia); with
polylinker; 1st strand cDNA was prepared from pooled bulk
breast tumor tissue, and was then primed with a Not I -
oligo(dT) primer. Double stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT7T3
vector. This library is the normalized version of
NCI_CGAP_Br1.1. Library was constructed by Bento Soares
and M. Fatima Bonaldo. "
BASE COUNT 88 a 76 c 104 g 89 t
ORIGIN

alignment scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2_COPY_38_46 x AA587454 ..

Align seq 1/1 to: AA587454 from: 1 to: 457

3 ArgArgAspTrpSerProSer 9
||||||||||||||||||||
330 CGCAGGGACTGGTCTCCTCT 350

seq_name: qb_gss20:AZ018678

seq_documentation_block:
LOCUS AZ018678 369 bp DNA GSS 25-FEB-2000
DEFINITION RPCI-23-334K19.TV RPCI-23 Mus musculus genomic clone RPCI-23-334K19
, DNA sequence.
ACCESSION AZ018678
VERSION AZ018678.1 GI:7094062
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 369)
AUTHORS Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinnet
,J., Lovins,M., Mcgann,S., Isoqaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished (1999)
COMMENT Other GSSs: RPCI-23-334K19.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter.dejong@med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderinginame.htm)
or from Research Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 334 row: K column: 19
Seq primer: T7
Class: BAC ends.
FEATURES Location/Qualifiers
source 1..369
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-334K19"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: kidney/Brain; Vector: pBACe3.6; Site 1:
[illegible]; Site 2: EcoRI; Insert [illegible]; Male
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies). "
BASE COUNT 106 a 100 c 73 g 82 t 8 others
ORIGIN

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0

```
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_38_46 x AZ018678   ..

 Align seq 1/1  to: AZ018678  from: 1  to: 369

        3 ArgArgAspIrpSerProSer 9
          ||||||||||||||||||||
      318 AGGAGAGACTGGAGTCCTTCA 338

seq_name: qb_est15:AI028200

seq_documentation_block:
LOCUS        AI028200      411 bp    mRNA            EST       27 APR 1998
DEFINITION  oy96b01.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1645129
            3' similar to contains Alu repetitive element;, mRNA sequence.
ACCESSION   AI028200
VERSION     AI028200.1  GI:3245906
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 411)
  AUTHORS   NCI-CGAP Email: cgapbs-r@mail.nih.gov
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
            , Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 874  Std Error: 0.00
            Seq primer: -40mlB twd. ET from Amersham
            High quality sequence stop: 398.
FEATURES             Location/Qualifiers
     source          1..411
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1645129"
                     /clone_lib="Soares_testis_NHT"
                     /sex="male"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech Laboratories
                     , Inc., and primed with  a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCAAIIIIIIIIIIIIIIIII 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       96 a     86 c    121 g    108 t
ORIGIN


alignment_scores:
            Quality:    7.00            Length:       7
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_38_46 x AI028200   ..

 Align seq 1/1  to: AI028200  from: 1  to: 411

        3 ArgArgAspIrpSerProSer 9
          ||||||||||||| ||||||||
      326 CGGAGGGACTGGTCCTTCT 346

seq_name: qb_est19:AI369207

seq_documentation_block:
LOCUS        AI369207      453 bp    mRNA            EST       16 FEB 1999
DEFINITION  qy75d06.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2017835 3',
            mRNA sequence.
ACCESSION   AI369207
VERSION     AI369207.1  GI:4147960
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 453)
  AUTHORS   NCI-CGAP Email: cgapbs-r@mail.nih.gov
  TITLE     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 494  Std Error: 0.00
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1..453
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2017835"
                     /clone_lib="NCI_CGAP_Brn25"
                     /tissue_type="anaplastic oligodendroglioma"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I  oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCAIAGGTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of  the modified pT7T3 vector.
                     Library is normalized, and was constructed  by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT       116 a    100 c     99 g    138 t
ORIGIN


alignment_scores:
            Quality:    7.00            Length:       7
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_38_46 x AI369207/rev   ..

 Align seq 1/1  to reverse of: AI369207  from: 1  to: 453

        3 ArgArgAspIrpSerProSer 9
          |||||||||||||||||||||
      127 AGGAGAGACTGGTCTCCAAGC 107

seq_name: qb_est33:BE827453
```

```
seq documentation block:
LOCUS        BF879453       484 bp    mRNA            EST       17 JAN 2001
DEFINITION   QV3 ET0171-241100 494-e08 ET0171 Homo sapiens cDNA, mRNA sequence.
ACCESSION    BF879453
VERSION      BF879453.1  GI:12269583
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 484)
  AUTHORS    Dias Neto,E., Garcia Correa,R., Verjovski Almeida,S., Briones,M.R.,
             Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
             Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
             Brumstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
             ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
             Simpson,A.J.
  TITLE      Shotgun sequencing of the human transcriptome with ORF expressed
             sequence tags
  JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE    20202663
COMMENT      Contact: Simpson A.J.G.
             Laboratory of Cancer Genetics
             Ludwig Institute for Cancer Research
             Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo SP,
             Brazil
             Tel: +55 11 2704922
             Fax: +55 11 2704707
             Email: asimpson@ludwig.org.br
             This sequence was derived from the FAPESP/LICR Human Cancer Genome
             Project. This entry can be seen in the following URL
             (http://www.ludwig.org.br/scripts/getxml2.pl?11 QV3&t2 QV3 ET0171
             241100 494-e08&t3 2000 11-24&t4 1)
             Seq primer: puc 18 forward
             High quality sequence start: 55
             High quality sequence stop: 484
FEATURES             Location/Qualifiers
     source          1..484
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone_lib "ET0171"
                     /dev_stage "Adult"
                     /note "Organ: lung_tumor; Vector: pucl8; Site_1: SmaI;
                     Site_2: SmaI; A mini library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector.  Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       91 a    152 c    105 g    136 t
ORIGIN


alignment scores:
            Quality:     7.00            Length:       7
              Ratio:  1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945b-2 copy 38-46 x BF879453    .

 Align seq 1/1  to: BF879453  from: 1  to: 484

        3 ArgArgAspIrpSerProSer 9
          ||||||||||||||||||||
      104 AGGAGGGATTGGTCACCGTCC 124

seq name: qb_est45:AW417078

seq documentation block:
LOCUS        AW417078       500 bp    mRNA            EST       24 JAN 2000
DEFINITION   xr26f02.x1 NCI_CGAP_Pr4 Homo sapiens cDNA clone IMAGE:2761275 3',
             mRNA sequence.
```

```
ACCESSION    AW417078
VERSION      AW417078.1  GI:6746344
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 500)
  AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
  JOURNAL    Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
             Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             www-bio.llnl.gov/bbrp/image/image.html
             Seq primer: -40UP from Gibco
             High quality sequence stop: 428.
FEATURES             Location/Qualifiers
     source          1..500
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone="IMAGE:2761275"
                     /clone_lib "NCI_CGAP_Pr4"
                     /tissue_type "Serous papillary carcinoma, high grade, 2
                     pooled tumors"
                     /lab_host "DH10B"
                     /note "Organ: uterus; Vector: pCMV SPORT6; Site 1: SalI;
                     Site 2: NotI; Cloned unidirectionally.  Primer: oligo dT.
                     Average insert size 1.48 Kb. Life Technologies catalog #:
                     11542-016"
BASE COUNT       113 a    112 c    154 g    121 t
ORIGIN


alignment scores:
            Quality:     7.00            Length:       7
              Ratio:  1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945b-2 copy 38-46 x AW417078    ..

 Align seq 1/1  to: AW417078  from: 1  to: 500

        3 ArgArgAspIrpSerProSer 9
          ||||||||||||||||||||
      324 CGCAGGGACTGGTCACCGTCT 344

seq name: qb_est70:BE105714

seq documentation block:
LOCUS        BE105714       519 bp    mRNA            EST       13-JUN-2000
DEFINITION   UI_R_Bn1_asi_1_09_0_UI.s1 UI_R_Bn1_Rattus norvegicus cDNA clone
             UI_R_Bn1_asi_1_09_0_UI 3', mRNA sequence.
ACCESSION    BE105714
VERSION      BE105714.1  GI:8497856
KEYWORDS     EST.
SOURCE       Norway rat.
  ORGANISM   Rattus norvegicus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae;
             Rattus.
REFERENCE    1  (bases 1 to 519)
  AUTHORS    Bonaldo,M.F., Lennon,G. and Soares,M.B
  TITLE      Normalization and subtraction: two approaches to facilitate gene
             discovery
```

```
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the Not I site
            and the oligo-dT track served to identify it as a clone from the
            normalized hippocampus library cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: clones will be available through
            Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA Yes.
FEATURES             Location/Qualifiers
     source          1..519
                     /organism "Rattus norvegicus"
                     /strain "Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-B0l-aju-c-05-0-UI"
                     /clone_lib "UI-R-B0l"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
                     /note "Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; the library
                     (UI-R-B0l) is a subtracted library derived from a mixture
                     of the following tissues: thalamus, cerebellum,
                     hypothalamus, medulla, pons, midbrain, cerebral cortex,
                     corpus striatum and hippocampus. For a detailed
                     description of the library from which this clone was
                     derived, please visit our web site at
                     ratest.enq.uiowa.edu. The subtraction has been previously
                     described in (Bonaldo, Lennon and Soares, Genome Research
                     6:791-806, 1996)
                     TAG_LIB-UI-R-B0l
                     TAG_TISSUE hippocampus
                     TAG_SEQ GATTG"
BASE COUNT       141 a    125 c    100 g    153 t
ORIGIN


alignment_scores:
            Quality:     7.00             Length:       7
              Ratio:    1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_38_46 x BE105714/rev    ..

 Align seq 1/1  to reverse of: BE105714  from: 1  to: 519

        8 ArgArgAspTrpSerProSer 9
          ||||||||||||||||||||||
      469 AGGAGAGACTGGTCCCCCTCG 449

seq_name: gb_est89:BF563188

seq_documentation_block:
LOCUS       BF563188       568 bp    mRNA            EST       12-DEC-2000
DEFINITION  UI-R-B0l-aju-c-05-0-UI.r1 UI-R-B0l Rattus norvegicus cDNA clone
            UI-R-B0l-aju-c-05-0-UI 5', mRNA sequence.
ACCESSION   BF563188
VERSION     BF563188.1  GI:11672919
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 568)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            clones will be available through Research Genetics (www.resgen.com)
            This clone is also available through the I.M.A.G.E. Consortium at
            LLNL (info@image.llnl.gov). IMAGE ID  1798144
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1..568
                     /organism "Rattus norvegicus"
                     /strain="Sprague Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-B0l-aju-c-05-0-UI"
                     /clone_lib "UI-R-B0l"
                     /dev_stage "adult"
                     /lab_host "DH10B (Life Technologies)"
                     /note "Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; the library
                     (UI-R-B0l) is a subtracted library derived from a mixture
                     of the following tissues: thalamus, cerebellum,
                     hypothalamus, medulla, pons, midbrain, cerebral cortex,
                     corpus striatum and hippocampus. For a detailed
                     description of the library from which this clone was
                     derived, please visit our web site at
                     ratest.enq.uiowa.edu. The subtraction has been previously
                     described in (Bonaldo, Lennon and Soares, Genome Research
                     6:791-806, 1996)"
BASE COUNT       156 a    137 c    153 g    122 t      1 others
ORIGIN


alignment_scores:
            Quality:     7.00             Length:       7
              Ratio:    1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_38_46 x BF563188    ..

 Align seq 1/1  to: BF563188  from: 1  to: 568

        8 ArgArgAspTrpSerProSer 9
          |||||||||||||||||||||
      491 AGGAGAGACTGGTCCCCCTCG 511

seq_name: gb_est85:BF285418

seq_documentation_block:
LOCUS       BF285418       576 bp    mRNA            EST       28-NOV-2000
DEFINITION  EST145009 Rat Gene Index, normalized rat, Rattus norvegicus cDNA
            Rattus norvegicus cDNA clone RGIFB03 3', sequence, mRNA sequence.
ACCESSION   BF285418
VERSION     BF285418.1  GI:11216578
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 576)
  AUTHORS   Malek,J.L., Cho,J., Lee,Y., Karamycheva,S., Parvizi,P., Cortes,G.,
```

```
            Sultana,R., Tsai,J., White,J., Quackenbush,J. and Lee,N.H.
  TITLE     Generation of ESTs from Normalized Rat Embryo, Bento Soares
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Lee, NH
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301) 838 4529
            Fax: (301)-838 0208
            Email: nhlee@tigr.org
            This clone is available through the ATCC, contact the ATCC
            tel#703 365 2700 for further information.
FEATURES             Location/Qualifiers
     source          1..576
                     /organism "Rattus norvegicus"
                     /db_xref "taxon:10116"
                     /clone "RGIFH04"
                     /clone_lib "Rat Gene Index, normalized rat, Rattus
                     norvegicus cDNA"
                     /tissue_type "mixed tissue"
                     /lab_host "DH5 alpha"
                     /note "Vector: pT3T7Pac; Site 1: EcoRI; Site 2: NotI;
                     Combination of ROV, RBR, REI, RLL, RPL, RLU, REM, RMU, RSP
                     , RHE, RPC, RFN"
BASE COUNT       160 a    148 c    118 g    150 t
ORIGIN


alignment scores:
            Quality:     7.00            Length:       7
              Ratio:    1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_38_46 x BF285418/rev   ..

 Align seq 1/1  to reverse of: BF285418  from: 1  to: 576

        4 ArgArgAspTrpSerProSer 9
          ||||||||||||||||||||||
      457 AGGAGACTGGTCCCCTCG 434
```

```
GM of: US-09-126-945B-2_COPY_46_54 to: GenEmbl:*   out_format : pls

Date: Sep 14, 2001  9:36 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL frame+_p2n.model -DEV-xlp
 -Q "/cgn2_6/ptodata/1/pubpct/US09126945B/runat_14.rge" -DB=GenEmbl -QFMT fastap -SUFFIX=olip2n.rge -GAPOP 4.500
 -GAPEXT 0.050 -MINMATCH=0.100 -LOOPCL 0.000 -LOOPEXT 0.000
 -QGAPOP=4.500 -QGAPEXT 0.050 -XGAPOP=60.000 -XGAPEXT 60.000
 -FGAPOP 6.000 -FGAPEXT 7.000 -YGAPOP=60.000 -YGAPEXT 60.000
 -DELOP 6.000 -DELEXT=7.000 -START=1 -MATRIX oligo
 -TRANS human40.cdi -LIST=45 -DOCALIGN 200 -THR_SCORE quality
 -THR MIN 1 -ALIGN 15 -MODE LOCAL -OUTFMT pir  NORM ext  MINGEN=
 -MAXLEN=2000000000  USER=US09126945B_*CGN1_1_7594 -NCPU 6
 -ICPU 3 -LONGLOG -NO_XLPXY -WAIT -THREADS 1

Search information block:
Query: US-09-126-945B-2_COPY_46_54
Query length: 9
Database: GenEmbl:*
Database sequences: 1344157
Database length: -856060004
Search time (sec): 6504.700000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000


score_list:
Sequence                 Strd Orig    ZScore    EScore Len  ! Documentation ..
qb_rol:AB019436           +    9.00   155.45      2.04  1704  ! AB019436 Mus musculus mRNA for
qb_pat1:AX048148          +    9.00   154.79      2.19  1894  ! AX048148 Sequence 1 from Paten
qb_pr4:AF071538           +    9.00   154.79      2.19  1894  ! AF071538 Homo sapiens Ets tran
qb_pr1:AB031549           +    9.00   154.78      2.20  1898  ! AB031549 Homo sapiens PSE mRNA
qb_pat1:AX048159          +    9.00   151.86      8.20  3817  ! AX048159 Sequence 12 from Pate
qb_pr6:AL157472           +    9.00   131.85     41.58  15666  ! AL157472 Human DNA sequence
qb_sts2:HS937A7           +    7.00   129.24     58.39   367  ! AL159275 H.sapiens STS from fi
qb_pl3:NIE9VPIFA          +    7.00   128.39     61.84   454  ! 725412 N.tabacum gene for malic
qb_pat1:A86202            -    7.00   125.60     62.72   634  ! A86202 Sequence 863 from Paten
qb_pat2:P62220            -    7.00   125.60     62.72   634  ! P62220 Sex determining region
qb_pr9:HSM801388          -    7.00   121.66    154.75  1203  ! AL133187 Homo sapiens mRNA, cD
qb_prl0:HUMNB1            +    7.00   121.65    154.80  1306  ! M58026 Human NB 1 mRNA, comple
qb_pr7:BC001044           -    7.00   120.29    163.13  1690  ! BC001044 Homo sapiens, Similar
qb_pr5:AK023986           -    7.00   119.72    167.15  1886  ! AK023986 Homo sapiens cDNA FLJ
qb_prl:AB055250           -    7.00   119.65    168.91  1911  ! AB055250 Macaca fascicularis b
qb_ba3:PPE6662            -    7.00   119.28    209.59  2051  ! AJ006662 Propionibacterium fr
qb_rol:AF134804           -    7.00   118.68    225.20  2490  ! AF134804 Mus musculus clone sk
qb_pr5:AF152484           -    7.00   118.43    232.61  2412  ! AF152484 Homo sapiens protease
qb_pr5:AF152486           -    7.00   118.36    234.74  2445  ! AF152486 Homo sapiens protease
qb_prl0:HUMCALPS          +    7.00   117.76    254.56  2749  ! M36707 Human calmodulin like p
qb_pr9:HSCAMPSG           +    7.00   117.75    254.80  2746  ! X13461 H.sapiens intronless ca
qb_pr5:AK001742           -    7.00   117.58    259.30  2885  ! AK001742 Homo sapiens cDNA FL
qb_pr5:AF152314           -    7.00   114.34    362.88  5260  ! AF152314 Homo sapiens protease
qb_pr5:AF152316           -    7.00   114.34    362.88  5260  ! AF152316 Homo sapiens protease
qb_bal:AE004647           +    7.00   110.51    642.63  10936  ! AE004647 Pseudomonas aeruginos
qb_bal:AE004271           -    7.00   110.39    652.39  11184  ! AE004271 Vibrio cholerae chro
qb_inl:AC006639           +    7.00   104.41   1.4e+03  35018  ! AC006639 Caenorhabditis elega
qb_in3:CELF43C11          +    7.00   104.41   1.4e+03  35018  ! AC017117 Caenorhabditis elega
qb_ba3:SVI011500          +    7.00   103.81   1.5e+03  39250  ! AJ011500 Streptomyces violace
qb_htq9:AC023865          -    7.00   101.10   2.1e+03  65824  ! AC023865 Homo sapiens chromos
qb_htq8:AC023223          -    7.00    99.65   2.6e+03  96742  ! AC023223 Homo sapiens chromos
qb_htq5:AC016815          -    7.00    99.06   2.8e+03  97110  ! AC016815 Homo sapiens chromos
qb_pr2:AC006262           +    7.00    98.58   3.0e+03  106328  ! AC006262 Homo sapiens chromo
qb_htq19:AF293336         +    7.00    98.46   3.0e+03  108986  ! AF293336 Homo sapiens chromo
qb_htq16:AC073744_1       +    7.00    98.41   3.0e+03  110000  ! Continuation (2 of 5) of AC0
qb_pr7:AP001065           -    7.00    98.14   3.1e+03  115798  ! AP001065 Homo Sapiens genomi
qb_pr2:AC008468           +    7.00    98.08   3.2e+03  117096  ! AC008468 Homo Sapiens chromo
qb_pr2:AC005481           +    7.00    98.00   3.2e+03  118802  ! AC005481 Homo Sapiens clone
qb_htq3:AC011348          +    7.00    97.99   3.2e+03  119100  ! AC011348 Homo Sapiens chromo
qb_pl2:AP000304           -    7.00    97.24   3.5e+03  137454  ! AP000304 Oryza sativa gen
qb_pr9:HSDJ977B1          -    7.00    96.90   3.7e+03  145068  ! AL050141 Human DNA Sequen
qb_pr2:AC007193           -    7.00    96.92   3.7e+03  146180  ! AC007193 Homo Sapiens chr
qb_htq2:AC010802          -    7.00    96.71   3.8e+03  152009  ! AC010802 Homo Sapiens clo
qb_htq24:AP000453         -    7.00    96.71   3.8e+03  152053  ! AP000453 Oryza Sativa chr
qb_htq10:AC025626         -    7.00    96.62   3.8e+03  154696  ! AC025626 Homo Sapiens clo

Seq_name: qb_rol:AB019436

Query of size 9 against: GenEmbl:*
LOCUS       AB019436     1704 bp    mRNA            ROD       26-JAN-2000
DEFINITION  Mus musculus mRNA for Prostate specific transcription factor ets,
            complete cds.
ACCESSION   AB019436
VERSION     AB019436.1  GI:4760537
KEYWORDS    Prostate specific transcription factor ets; mPse.
SOURCE      Mus musculus (strain:C57BL/6) Adult male Colon Epithelial cDNA to
            mRNA.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Yamada,N., Iamai,Y., Miyamoto,H. and Nozaki,M.
  TITLE     Cloning and expression of the mouse Pse gene encoding a novel Ets
            family member
  JOURNAL   Gene 241 (2), 267-274 (2000)
  MEDLINE   20137504
REFERENCE   2  (bases 1 to 1704)
  AUTHORS   Nozaki,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-NOV-1998) to the DDBJ/EMBL/GenBank databases. Masami
            Nozaki, Osaka University, Institute for Microbial Diseases; 3-1,
            Yamadaoka, Suita, Osaka 565-0871, Japan
            (E-mail:mnozaki@biken.osaka-u.ac.jp, Tel:81-6-879-8338,
            Fax:81-6-879-8339)
FEATURES             Location/Qualifiers
     source          1..1704
                     /organism="Mus musculus"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /cell_type="Epithelial"
                     /dev_stage="Adult"
                     /sex="male"
                     /tissue_type="Colon"
     gene            362..1339
                     /gene="mPse"
     CDS             362..1339
                     /gene="mPse"
                     /codon_start=1
                     /product="Prostate specific transcription factor ets"
                     /protein_id="BAA77329.1"
                     /db_xref="GI:4760537"
                     /translation="MGSAsrGLsNVSPsrTLLEPDVAcKrGLENAAScAMGFENQEWS
                     PSPDATFFQGLSAFYLSYENMYPDDSSWVAKVPEARAYEDRPEEPPEgCPVTDSQASGS
                     TLDEHSLDQVQSMVVGEVLKDIETACKLLNITADPGDWSPGNVQKWLLWIEHQYKLPP
                     ACKAFQFLGCKFLCAMSEFQEPQPAPLGCDVLHAHLLCWFSAAWMKEPISPCTLHYCA
                     SISEEGWHPGEVDSSCSGQPIHLWQFLKELLIEPESYCPFIRWLNEENSKNIFEIPDSAQ
                     VAFFWYGENFLAMNYQKSSFSIFLYPVQLIRKPDISQPLVYQFVHPV"
     polyA_site      1704
                     /note="10 a nucleotides"
BASE COUNT       369 a    527 c    491 g    317 t
ORIGIN


Alignment_scores:
            Quality:    9.00              Length:      9
              Ratio:   1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment_block:
 US-09-126-945B-2_COPY_46_54 x AB019436  ..

 Align seq 1/1  to: AB019436  from: 1  to: 1704
```

```
        1 SerProProAlaThrProGluGlnGly 9
          ||||||||||||||||||||||||||||
      497 AGTCCACCCGCCACCCCTGAGCAGGGC 523

seq_name: qb_pat1:AX048148

seq documentation block:
LOCUS       AX048148      1894 bp    DNA             PAT       15-DEC-2000
DEFINITION  Sequence 1 from Patent WO0078892.
ACCESSION   AX048148
VERSION     AX048148.1  GI:11876967
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1894)
  AUTHORS   Kaufmann,J. and Xin,H.
  TITLE     Expression of ets domain proteins in cancer
  JOURNAL   Patent: WO 0078892 A 1 28-DEC-2000;
            Chiron Corporation (US)
FEATURES             Location/Qualifiers
     source          1..1894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             416..1423
                     /note="Human JKETS."
                     /codon_start=1
                     /protein_id="CAC19092.1"
                     /db_xref="GI:11876968"
                     /translation="MQSASPQLSSVSPSHLLLPPPELVSPTGLEPKAAACAVCLERREWS
                     PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAPCASSREEPPEEPEQCPVTDSQAPA
                     GSLDLVPCGLTLEEHSLEQVQSMVVCEVLKDIETACKLLNLTADEMDWSPSNVQKWLL
                     WTEHQYRLPPMGKAFQELAGKELCAMSEEQFRQRSPLGGDVLHAHLDIWKSAAWMKER
                     TSPGATHYCASISEESWTDSEVDSSCSGQPTRLWQFLKELLLKPHSYGRFIRWLNKEK
                     GIFKIEDSAQVARLWGIRKNKPAMNYDKLSRSLRQYYKKGIIRKPDISQRLVYQFVHP
                     I"
BASE COUNT       368 a    654 c    571 g    302 t
ORIGIN


alignment scores:
            Quality:     9.00              Length:       9
              Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2_COPY_46_54 x AX048148  .

 Align seq 1/1  to: AX048148  from: 1  to: 1894

        1 SerProProAlaThrProGluGlnGly 9
          ||||||||||||||||||||||||||||
      551 AGTCCACCCGCCACCCCTGAGCAGGGC 577

seq_name: qb_pr4:AF071538

seq documentation block:
LOCUS       AF071538      1894 bp    mRNA            PRI       10-JAN-2000
DEFINITION  Homo sapiens Ets transcription factor PDEF (PDEF) mRNA, complete
            cds.
ACCESSION   AF071538
VERSION     AF071538.1  GI:4007417
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1894)
  AUTHORS   Oettgen,P., Finger,E., Sun,Z., Akbarali,Y., Thamrongsak,U.,
            Boltax,J., Grall,F., Dube,A., Weiss,A., Brown,L., Quinn,G., Kas,K.,
            Endress,G., Kunsch,C. and Libermann,T.A.
  TITLE     PDEF, a novel prostate epithelium specific ets transcription
            factor, interacts with the androgen receptor and activates
            prostate-specific antigen gene expression
  JOURNAL   J. Biol. Chem. 275 (4), 1216-1225 (2000)
  MEDLINE   20092894
REFERENCE   2  (bases 1 to 1894)
  AUTHORS   Libermann,T.A., Oettgen,P., Kunsch,C. and Finger,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUN-1998) Medicine, Beth Israel Deaconess Medical
            Center, 330 Brookline Ave., Boston, MA 02215, USA
FEATURES             Location/Qualifiers
     source          1..1894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6p21.3"
                     /tissue_type="prostate"
     gene            1..1894
                     /gene="PDEF"
     CDS             416..1423
                     /gene="PDEF"
                     /note="prostate epithelium-specific"
                     /codon_start=1
                     /product="Ets transcription factor PDEF"
                     /protein_id="AAC95296.1"
                     /db_xref="GI:4007418"
                     /translation="MQSASPQLSSVSPSHLLLPPPELVSPTGLEPKAAACAVCLERREWS
                     PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAPCASSREEPPEEPEQCPVTDSQAPA
                     GSLDLVPCGLTLEEHSLEQVQSMVVCEVLKDIETACKLLNLTADEMDWSPSNVQKWLL
                     WTEHQYRLPPMGKAFQELAGKELCAMSEEQFRQRSPLGGDVLHAHLDIWKSAAWMKER
                     TSPGATHYCASISEESWTDSEVDSSCSGQPTRLWQFLKELLLKPHSYGRFIRWLNKEK
                     GIFKIEDSAQVARLWGIRKNKPAMNYDKLSRSLRQYYKKGIIRKPDISQRLVYQFVHP
                     I"
BASE COUNT       368 a    654 c    571 g    302 t
ORIGIN


alignment scores:
            Quality:     9.00              Length:       9
              Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2_COPY_46_54 x AF071538  .

 Align seq 1/1  to: AF071538  from: 1  to: 1894

        1 SerProProAlaThrProGluGlnGly 9
          ||||||||||||||||||||||||||||
      551 AGTCCACCCGCCACCCCTGAGCAGGGC 577

seq_name: qb_pr1:AB031549

seq documentation block:
LOCUS       AB031549      1898 bp    mRNA            PRI       20-JAN-2000
DEFINITION  Homo sapiens PSE mRNA for prostate ets, complete cds.
ACCESSION   AB031549
VERSION     AB031549.1  GI:6721497
KEYWORDS    prostate ets.
SOURCE      Homo sapiens adult male prostate epithelium prostate cancer
            cell, line:JCA-3 cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Yamada,N., Iwai,Y., Miyamoto,H. and Nozaki,M.
  TITLE     Cloning and expression of the mouse Pse gene encoding a novel Ets
            family member
  JOURNAL   Gene 241 (2), 267-274 (2000)
  MEDLINE   20137504
REFERENCE   2  (bases 1 to 1898)
  AUTHORS   Nozaki,M.
  TITLE     Direct Submission
```

```
  JOURNAL    Submitted (25-AUG-1999) to the EMBL/GenBank/DDBJ databases. Masaki
             Nozaki, Osaka University, Institute for Microbial Diseases; 2-1,
             Yamadaoka, Suita, Osaka 565-0871, Japan
             (E-mail:mnozaki@biken.osaka-u.ac.jp, Tel:81-6-879-8279,
             Fax:81 6 879 8339)
FEATURES             Location/Qualifiers
     source          1. .1898
                     /organism="H. sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="PC-3"
                     /cell_type="prostate carcinoma"
                     /chromosome="6"
                     /dev_stage="adult"
                     /map="6p21.2-21.3"
                     /sex="male"
                     /tissue_type="prostate epithelium"
                     /note="vector:pUC18"
     gene            406. .1413
                     /gene="PSE"
     CDS             406. .1413
                     /gene="PSE"
                     /codon_start=1
                     /product="prostate ets"
                     /protein_id="BAA89543.1"
                     /db_xref="GI:6721498"
                     /translation="MGSASPGLSSVSPSHLLPPDTVSKQLEKAAAGAVGLERRDWS
                     PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAPGASSREEPPEEPEQCPVIDSQAPA
                     GSLDLVPSGLTLEFHSLEQVQSMVVQEVLKQFTACFLLNLIADFMDWSFSNVQEWLL
                     WIEHQYKLEPMQKAFQELAGKELTAMSFEQLQESFLGSQVELHARLLDWKSAAWMEER
                     TSPGAIHYCASISEESWTDSEVDSSCSSQADTHIWQFLGELQKLGDRSYCPKTIWELNFAK
                     GIEKIEDSAQVARLWGIRKNKPAMNYDKLSRSIRQYYSKQKILRKPIDISQRLVYQFVHP
                     1"
BASE COUNT       380 a    651 c    567 g    300 t
ORIGIN


alignment_scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_46_54 x AB031549  ..

 Align seq 1/1  to: AB031549  from: 1  to: 1898

       1 SerProProAlaThrProGluGlnGly 9
         || |||||||||||||||||||||||||
     541 AGTCCACCCGCCACGGCCCGAGCAGGGC 567

seq_name: gb_pat1:AX048159

seq_documentation_block:
LOCUS        AX048159      3317 bp    DNA              PAT       15-DEC-2000
DEFINITION  Sequence 12 from Patent WO0070092
ACCESSION   AX048159
VERSION     AX048159.1  GI:11876975
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3317)
  AUTHORS   Kaufmann,J. and Xin,H.
  TITLE     Expression of ets-domain proteins in cancer
  JOURNAL   Patent: WO 9070092-A 12 23-NOV 2000;
            Chiron Corporation (US)
FEATURES             Location/Qualifiers
     source          1. .3317
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="CSEF-encoding sequence with promoter (Figs. 2A-2B)"
     promoter        1. .1380
                     /to="promoter"
     CDS             1015. .2042
                     /note="CSEF coding sequence"
                     /codon_start=1
                     /protein_id="CAC19004.1"
                     /db_xref="GI:11876976"
                     /translation="MGSASPGLSSVSPSHLLPPDTVSKQLEKAAAGAVGLERRDWS
                     PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAPGASSREEPPEEPEQCPVIDSQAPA
                     GSLDLVPSGLTLEFHSLEQVQSMVVQEVLKQFTACFLLNLIADFMDWSFSNVQEWLL
                     WIEHQYKLEPMQKAFQELAGKELTAMSFEQLQESFLGSQVELHARLLDWKSAAWMEER
                     TSPGAIHYCASISEESWTDSEVDSSCSSQADTHIWQFLGELQKLGDRSYCPKTIWELNFAK
                     GIEKIEDSAQVARLWGIRKNKPAMNYDKLSRSIRQYYSKQKILRKPIDISQRLVYQFVHP
                     1"
BASE COUNT       710 a    1026 c    970 g    611 t
ORIGIN


alignment_scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_46_54 x AX048159  ..

 Align seq 1/1  to: AX048159  from: 1  to: 3317

       1 SerProProAlaThrProGluGlnGly 9
         || |||||||||||||||||||||||||
    1950 AGTCCACCCGCCACGGCCCGAGCAGGGC 1976

seq_name: gb_pri6:AL157372

seq_documentation_block:
LOCUS        AL157372    150660 bp    DNA              PRI       29-NOV-2000
DEFINITION  Human DNA sequence from clone RP11-475E1 on chromosome 6, complete
            sequence.
ACCESSION   AL157372
VERSION     AL157372.18  GI:11493236
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 150660)
  AUTHORS   Skuce,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 30, 2000 this sequence version replaced gi:11342772.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows.  An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements.  Where the sequence is
            ambiguous, there is an annotation using the 'unsure' feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            EMBL: EMBL; SW:, SWISSPROT; TR:, TREMBL; WP:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/; pg 1%: ~ pc e
            was generated from part of bacterial clone contig of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-475E1 is from the library RPCI-11.2 constructed at the Roswell
            Park Cancer Institute by the group of Pieter de Jong. For further
```

```
            details see http://barpac.med.buffalo.edu/
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RPI1 475E1 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RPI1 475E1 is at 1 in this sequence. The
            true left end of clone RP3 391022 is at 150561 in this sequence.
            The true right end of clone RPI1 302L10 is at 114044 in this
            sequence.
FEATURES             Location/Qualifiers
     source          1..150660
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RPI1 475E1"
                     /clone_lib="RPCI-11.2"
BASE COUNT    45470 a   38887 c   38740 g   27563 t
ORIGIN


alignment scores:
            Quality:     9.00            Length:       9
              Ratio:     1.000             Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_46_54 x AL157472/rev   ..

 Align seq 1/1  to reverse of: AL157472  from: 1  to: 150660

       1 SerProProAlaThrProGluGlnGly 9
         |||||||||||||||||||||||||||
  129605 AGTCCACCCGCCACGCCCGAGCAGGGC 129579

seq name: qb_sts2:HSH27A7

seq documentation block:
LOCUS        HSH27A7         307 bp    DNA             STS       06-MAR-2000
DEFINITION   H.sapiens STS from flow sorted chromosome 13 random shear fragment,
             sequence tagged site.
ACCESSION    AL159075
VERSION      AL159075.1  GI:7208864
KEYWORDS     STS.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 307)
  AUTHORS    Hunt,S., Sims,S., Willey,D., Carter,N. and Ross,M.
  TITLE      Direct Submission
  JOURNAL    Submitted (06-MAR-2000) Email contact: humpory@hgmp.mrc.ac.uk
COMMENT      Marker sts-hr-13_27A7 (Primer A : CAAGTTTGCATTCAACCCC; Primer B :
             GAGGTGAGTGGGTCAAAATCA; amplimer size : 151 bp) was developed from a
             single pass sequencing read from H.sapiens flow sorted chromosome
             13 random shear fragment, SC14pH27a7. Vector : pUC18 Site : SmaI
             Further information : http://www.hgmp.mrc.ac.uk/GenomeData/.
FEATURES             Location/Qualifiers
     source          1..307
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /sex="female"
                     /clone="SC14pH27a7"
                     /tissue_type="Lymphoblastoid cell line"
                     /clone_lib="SC14pH"
BASE COUNT       102 a      67 c      60 g      77 t       1 others
ORIGIN


alignment scores:
            Quality:     7.00            Length:       7
              Ratio:     1.000             Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_46_54 x HSH27A7   ..

 Align seq 1/1  to: HSH27A7  from: 1  to: 307

       2 ProProAlaThrProGlnGln 8
         |||||||||||||||||||||
     160 CCAGCTGCGACCCCAGACAG 180

seq name: qb_pl3:NIE9VPIIA

seq documentation block:
LOCUS        NIE9VPIIA       354 bp    DNA             PLN       09-FEB-1996
DEFINITION   N.tabacum gene for maize transcriptional activator Vp1 homolog
             (partial).
ACCESSION    Z26312
VERSION      Z26312.1  GI:402233
KEYWORDS     transcriptional activator.
SOURCE       common tobacco.
  ORGANISM   Nicotiana tabacum
             Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
             Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
             I; Solanales; Solanaceae; Nicotiana.
REFERENCE    1  (bases 1 to 354)
  AUTHORS    Phillips,J.P.
  TITLE      Direct Submission
  JOURNAL    Submitted (15-SEP-1994) Phillips J. P., Institute of Plant Genetics
             and Crop Plant Research, Molecular Cell Biology, Corrensstrasse 3,
             D-06466 Gatersleben, Germany
REFERENCE    2  (bases 1 to 354)
  AUTHORS    Phillips,J. and Conrad,U.
  TITLE      Genomic sequences from Nicotiana tabacum homologous to the maize
             transcriptional activator gene Viviparous 1
  JOURNAL    J. Plant Physiol. 144, 760-761 (1994)
FEATURES             Location/Qualifiers
     source          1..354
                     /organism="Nicotiana tabacum"
                     /strain="Xanthi nc"
                     /db_xref="taxon:4097"
                     /clone="EMBL3 Clone 133, subclone E9"
                     /dev_stage="Seedling"
                     /clone_lib="EMBL3 from Clontech"
     CDS             <1..>112
                     /note="Open reading frame is homologous to the maize
                     transcriptional activator Vp1."
                     /citation=[2]
                     /codon_start=2
                     /protein_id="CAA81222.1"
                     /db_xref="GI:402234"
                     /db_xref="SPTREMBL:Q40498"
                     /translation="PTPTSISSPSSSIVDDQQLQSSSISIHQQLESVVHG"
BASE COUNT        81 a     148 c      65 g      60 t
ORIGIN


alignment scores:
            Quality:     7.00            Length:       7
              Ratio:     1.000             Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_46_54 x NIE9VPIIA   ..

 Align seq 1/1  to: NIE9VPIIA  from: 1  to: 354

       1 SerProProAlaThrProGlu 7
         |||||||||||||||||||||
     180 TCACCCCAGCCACGCCCGAA 200

seq name: qb_pat1:AR86202

seq documentation block:
```

```
LOCUS       A86202          614 bp    DNA             PAT       21-JAN-2000
DEFINITION  Sequence 861 from Patent EP0866129.
ACCESSION   A86202
VERSION     A86202.1  GI:6734801
KEYWORDS    .
SOURCE      Eremothecium gossypii.
  ORGANISM  Eremothecium gossypii
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Eremotheciaceae; Eremothecium.
REFERENCE   1  (bases 1 to 614)
  AUTHORS   Mohr,C. and Knechtle,P.
  TITLE     Genomic DNA sequences of Ashbya gossypii and uses thereof
  JOURNAL   Patent: EP 0866129-A 861 23-SEP-1998;
            CIBA GEIGY AG (CH)
FEATURES             Location/Qualifiers
     source          1..614
                     /organism="Eremothecium gossypii"
                     /db_xref="taxon:33169"
BASE COUNT       91 a    163 c    281 g     79 t
ORIGIN


alignment_scores:
             Quality:    7.00              Length:       7
               Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_46_54 x A86202/rev    .

 Align seq 1/1  to reverse of: A86202  from: 1  to: 614

       1 SerProProAlaThrProGln 7
         ||||||||||||||||||||
     143 TCGCCGCCCGCCACCCCGCAG 123

seq_name: gb_pat2:E66220

seq_documentation_block:
LOCUS       E66220          614 bp    DNA             PAT       07-FEB-2001
DEFINITION  Genome DNA sequence of Ashbya gossypii and uses thereof.
ACCESSION   E66220
VERSION     E66220.1  GI:13023561
KEYWORDS    "P 1999225770-A/860
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 614)
  AUTHORS   Peter P.P.P.P., Steiner C.M., Hermann,W.P.F.F. and Robishaw.
  TITLE     Genome DNA sequence of Ashbya gossypii and uses thereof
  JOURNAL   Patent: JP 1999225770-A 860 24-AUG-1999;
            NOVARTIS AG
COMMENT     OS   PAG1542RP
            PN   JP 1999225770-A/860
            PD   24-AUG-1999
            PF   05-JAN-1998 JP 1998076818
            PR   31-DEC-1996 CH        16/97
            PI   PETER PHILLIPSEN,REINER PELHEMANN,SABINE STEINER,CHRISTINE
            MORE, PI   JUERGEN WENDLAND,PHILLIP KUNDHIIT?RU,CORINNE KURISHUN P?
              C12N15/09 C12N1/15 C12Q 01 C12Q 68 C12N15/ C12Q 64 P?
            (C12N15/09 C12R1:865) (C12N1/15 C12R1:645)
            (C12N15/80,C12R1:645).
            PC   (C12N15/80 C12R1:865)
            CC   Strandedness: Single.
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers
            FT   source          1..614
            FT                   /organism="PAG1542RP".
FEATURES             Location/Qualifiers
     source          1..614
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       91 a    163 c    281 g     79 t
ORIGIN


alignment_scores:
             Quality:    7.00              Length:       7
               Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_46_54 x E66220/rev    .

 Align seq 1/1  to reverse of: E66220  from: 1  to: 614

       1 SerProProAlaThrProGln 7
         ||||||||||||||||||||
     143 TCGCCGCCCGCCACCCCGCAG 123

seq_name: gb_pri9:HSM801488

seq_documentation_block:
LOCUS       HSM801488      1303 bp    mRNA            PRI       18-FEB-2000
DEFINITION  Homo sapiens mRNA; cDNA DKFZp434G0522 (from clone DKFZp434G0522).
ACCESSION   AL133107
VERSION     AL133107.1  GI:6458592
KEYWORDS
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1303)
  AUTHORS   Blum,H., Bannasch,S., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-1999) MIPS, Am Klopferspitz 18a, D-82152
            Martinsried, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
            sequenced by LMU (Ludwig Maximilians University,
            Munich/Germany) within the cDNA sequencing consortium of the German
            Genome Project.
            This clone (DKFZp434G0522) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://www.mips.biochem.mpg.de/proj/human_cDNA/.
FEATURES             Location/Qualifiers
     source          1..1303
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZp434G0522"
                     /clone_lib="434 (synonym: hbts4); Vector pSport1; host
                     DH10B; sites Not I + Sal I"
                     /dev_stage="adult"
                     /tissue_type="testis"
     gene            409..1185
                     /gene="DKFZp434G0522"
     CDS             409..1185
                     /gene="DKFZp434G0522"
                     /note="similarity to Kelsp and dictyostelium rab protein"
                     /codon_start=1
                     /p                  protein"
                     /protein_id="CAB03413.1"
                     /db_xref="GI:6458593"
                     /translation="MVAWKFQLILLGDPHLSKGRKVVVNGVYAFNQFDSLSPEVVK
                     RMNFSGVKFTPKSGFSVAMAFNHQETLFFGGVGLDIRDFESLSGEFENLIYEYLAIKNR
                     WILQKPELESGWKFTRKEIVREGRAAFEVGSLTPIAIEPAVREFIEKPVQPFLACAVP
                     ISAQDPAVLSEAIGASVRPDAQERAEGGWPRPGCELSPMV"
     polyA_signal    1206..1270
     polyA_site      1279
BASE COUNT      292 a    322 c    423 g    261 t
ORIGIN
```

```
alignment scores:
            Quality:     7.00              Length:       7
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_46_54 x HSM801388/rev   ..

 Align seq 1/1  to reverse of: HSM801388  from: 1  to: 1303

        1 SerProProAlaThrProGlu 7
          |||||||||||||||||||||
      933 TCTCCTCCAGCCACTCCTGAG 913

seq name: qb pri0:HUMNB1

seq documentation block:
LOCUS        HUMNB1       1306 bp    mRNA            PRI       27 APR 1993
DEFINITION  Human NB 1 mRNA, complete cds.
ACCESSION   M58026
VERSION     M58026.1  GI:189080
KEYWORDS    NB 1.
SOURCE      Human breast epithelial cell, cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1306)
  AUTHORS   Yaswen,P., Smell,A., Peehl,D.M., Trask,D.K., Saper,R. and
            Stampfer,M.R.
  TITLE     Down-regulation of a calmodulin-related gene during transformation
            of human mammary epithelial cells
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 87, 7360-7364 (1990)
  MEDLINE   91017505
FEATURES             Location/Qualifiers
     source          1..1306
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     mRNA            <1..575
                     /gene="NB 1"
     gene            1..575
                     /gene="NB 1"
     CDS             126..575
                     /gene="NB 1"
                     /codon_start=1
                     /protein_id="AAA46456.1"
                     /db_xref="GI:189081"
                     /translation="MAT[illegible]TEAELRDMMSEIDRDGNGTVDFPEFLGMMARKMKDTDSEEELIEAFRVFDKDGNGFVS
                     AAELRHVMTRLGEKLSDEEVDEMIRAADIDGDGQVNYEEFVRVLVSK"
BASE COUNT       258 a    434 c    403 g    211 t
ORIGIN


alignment scores:
            Quality:     7.00              Length:       7
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_46_54 x HUMNB1   ..

 Align seq 1/1  to: HUMNB1  from: 1  to: 1306

        1 SerProProAlaThrProGlu 7
          |||||||||||||||||||||
       26 TCTCCACCTGGCCACCCCAGAG 46

seq name: qb pri7:BC001044

seq documentation block:
LOCUS        BC001044     1690 bp    mRNA            PRI       16-MAR-2001
DEFINITION  Homo sapiens, Similar to hypothetical protein DKFZp434G0522, clone
            MGC:1180, mRNA, complete cds.
ACCESSION   BC001044
VERSION     BC001044.1  GI:12654436
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1690)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-DEC-2000) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 41 Center Drive, Room 1A03, Bethesda, MD 20892-2590,
            USA
  REMARK    [illegible]
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            Contact: amadan@systemsbiology.org
            Amy Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
            Greene, Mark Ketteman and Anuradha Madan

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 7 Row: b Column: 5.
FEATURES             Location/Qualifiers
     source          1..1690
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="MGC:1180 IMAGE:4505126"
                     /tissue_type="Placenta, choriocarcinoma"
                     /clone_lib="NIH MGC_21"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     CDS             19..1410
                     /codon_start=1
                     /product="Similar to hypothetical protein DKFZp434G0522"
                     /protein_id="AAH01044.1"
                     /db_xref="GI:12654437"
                     /translation="MGLKCLELEECPALELAAEMEEEVCEESPELITTLYNELYVYN
                     IKKDIWIKVDTPSPPRRCAHQAVVVPQQGQLWVFGGEFASPNGRQLYYFDFKDLWVLH
                     [illegible]WSELSFSGTGPTPESCQGMSVTPQGGLVVYGGYSPQFVEELQVLKSTDGSIMELLEKFED
                     GREDKWVWTRMNPSGVKPTPRSGFSVAMAPNHQTLFFGGVCDEERREESLSGEFFNDLY
                     FLDATRNRWFEAQGLKGPKSEKKKRRKGKKEEPEGGSRPAVGGAGTQGPVQLVKFVVAE
                     DGTVVTIKQVLTAPGSAGQPKSEDEDSLEEAGSPAPGPCPKSNAMLAVKHQVLAVYGG
                     MFEAGDRQVTLSDLHCLDLHRMEAWKALVEMDPETQEWLEETDSEEDSEEVEGAEGGV
                     DDEDSGEESGAED"
BASE COUNT       399 a    424 c    579 g    288 t
ORIGIN


alignment scores:
            Quality:     7.00              Length:       7
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_46_54 x BC001044/rev   ..

 Align seq 1/1  to reverse of: BC001044  from: 1  to: 1690

        1 SerProProAlaThrProGlu 7
          |||||||||||||||||||||
     1418 TCTCCTCCAGCCACTCCTGAG 1298

seq name: qb pri5:AK023986
```

```
seq_documentation_block:
LOCUS        AK023986      1886 bp    mRNA            PRI       29-SEP-2000
DEFINITION   Homo sapiens cDNA FLJ13924 fis, clone Y79AA1000540, weakly similar
             to CELL POLARITY PROTEIN TEA1.
ACCESSION    AK023986
VERSION      AK023986.1  GI:10436212
KEYWORDS     oligo capping; fis (full insert sequence).
SOURCE       Homo sapiens retinoblastoma cell_line:Y79 cDNA to mRNA.
             clone_lib:Y79AA1 clone:Y79AA1000540.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (sites)
  AUTHORS    Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
             Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
             Wakatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
             Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
             Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Muraoka,M.,
             Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
             Nakamura,Y., Nanahara,K., Masuho,Y., Nitamiya,K. and Iwayanagi,T.
  TITLE      NEDO human cDNA sequencing project
  JOURNAL    Unpublished (2000)
REFERENCE    2  (bases 1 to 1886)
  AUTHORS    Isogai,T. and Otsuki,T.
  TITLE      Direct Submission
  JOURNAL    Submitted (23 APR 2000) to the EMBL/GenBank/DDBJ databases.
             Isogai, Helix Research Institute, Genomics Laboratory; 1582-3 Yana,
             Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,
             Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT      NEDO human cDNA sequencing project supported by Ministry of
             International Trade and Industry of Japan; cDNA full insert
             sequencing: Research Association for Biotechnology; cDNA library
             construction, 5'- & 3'-end one pass sequencing and clone selection,
             Helix Research Institute (supported by Japan Key Technology Center
             etc.) and Department of Virology, Institute of Medical Science,
             University of Tokyo.
FEATURES             Location/Qualifiers
     source          1..1886
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="Y79"
                     /cell_type="retinoblastoma"
                     /clone="Y79AA1000540"
                     /clone_lib="Y79AA1"
                     /note="cloning vector: pME18SFL3"
     CDS             60..1622
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAB14756.1"
                     /db_xref="GI:10436213"
                     /translation="MKEPFEEPKPWPAGKIAAPMGPVPVPPPPPPPPPPAIAHPG
                     TLDAPETQTVEIPPPPPSTKPNASLPVPPPPTPTTPPPNLPTPPPPPPLYVPNG
                     RKDTWPKVDIPSPPPPPPCAHQAVVVPQQQQLWVPPQCEPAGPNCPLPYPYPDLWVLHL
                     ATKTWPEQVKSLGGGSGPSCPRMVAWRQLILPSCTHPSTPLYIYYNLVYAFNLLTFTW
                     SKLSPSGTCPIPPSCCQMSVIPQYCIVVPQYVPQQAVPPLVLPCLPPPCCMPLLKPPLC
                     REDKWVWTRMNPSCVKPTIPPSCPSVAMAPNHQTLLPCVCIPPPPPSLSCCPPPNCLYI
                     YDATRNPWPPCQLPCPESEPPPPRPCRPPPPPCCSPQAPCCPAPCCPVLIVPPVVAEI
                     GTVVTIKQVLTAPGSAGQPRSPEDPDSLPEAGSPAPSPCPPRCNAMLAVKHGVLYVYGGM
                     PEAGDPQVTLSDLHCLDLDTRMWAWKALVPMPPPLTWLPTTLTPTPCPPVPCPAPCVP
                     DEDSGEESGAED"
BASE COUNT       428 a    484 c    633 g    341 t
ORIGIN


alignment_scores:
            Quality:     7.00            Length:       7
              Ratio:    1.000              Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_copy_46_54 x AK023986/3'-...  ..

 Align seq 1/1  to reverse of: AK023986  from: 1  to: 1886
```

```
        1 Seq1: AtoAlaLeuProGlu 7
          ||||:|||||||||||||||||
     1530 TCTCCTCCAGCCACTCCACAG 1510

Seq_name: gb_pri:AB055250

Seq_documentation_block:
LOCUS        AB055250      1911 bp    mRNA            PRI       07-FEB-2001
DEFINITION   Macaca fascicularis brain cDNA, clone:QtlA-10033.
ACCESSION    AB055250
VERSION      AB055250.1  GI:12698094
KEYWORDS     fis (full insert sequence).
SOURCE       Macaca fascicularis adult male frontal lobe left cDNA to mRNA.
             clone_lib:macaque brain cDNA library QtlA clone:QtlA-10033.
  ORGANISM   Macaca fascicularis
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
             Cercopithecinae; Macaca.
REFERENCE    1  (sites)
  AUTHORS    Osada,N., Hida,M., Kusuda,J., Tanuma,R., Isoki,K., Hirai,M.,
             Terao,K., Suzuki,Y., Sugano,S. and Hashimoto,K.
  TITLE      Isolation of full-length cDNA clones from macaque brain cDNA
             libraries
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 1911)
  AUTHORS    Hashimoto,K., Osada,N., Hida,M., Kusuda,J. and Sugano,S.
  TITLE      Direct Submission
  JOURNAL    Submitted (02 FEB 2001) Katsuyuki Hashimoto, National Institute of
             Infectious Diseases, Division of Genetic Resources, 23-1, Toyama
             1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
             (E-mail:k-hasimo@nih.go.jp, URL:http://www.nih.go.jp/~hasimo/,
             Tel:81-3-5285-1111(ex.2120), Fax:81-3-5285-1181)
COMMENT      Lab host:    TOP10
             Vector:      pME18S-FL3 (Acc.No. AB009864)
             R. Site1:    DraIII (CACTGTGTG)
             R. Site2:    DraIII (CACAGTGTG)
             Description: 1st strand cDNA was primed with an oligo(dT) primer
             [AGTGGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was synthesized
             using specific 5' and 3' primers and amplified by PCR. The PCR
             product was digested with SfiI and size selection was performed to
             exclude fragments <1.5kb.The SfiI-digested PCR product was cloned
             into distinct DraIII sites of pME18S-FL3. XhoI sites just outside
             the DraIII sites can be used to isolate the cDNA insert. Libraries
             were constructed by Sugano et al.(University of Tokyo, Institute of
             Medical Science). Custom primer used for sequencing:
             (5' end primer [CTTCTGCTCTAAAAGCTGCG];
             3' end primer [CGACGTGCAGCTCGAGCACA] ).
FEATURES             location/Qualifiers
     source          1..1911
                     /organism="Macaca fascicularis"
                     /db_xref="taxon:9541"
                     /clone="QtlA-10033"
                     /sex="male"
                     /tissue_type="frontal lobe left"
                     /clone_lib="Macaque brain cDNA library, QtlA"
                     /dev_stage="adult"
     CDS             68..1573
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="BAB21874.1"
                     /db_xref="GI:12698095"
                     /translation="MKEKGEEEKKEEEEEEEEALTAAPGPLPVAPPQPVPLPPPPPSPEL
                     NASLSVHDPEKDELTLPGGEYPNGQKTPLYNPLYVVYNPKNAWTPVPLPSPPPPPCAHQ
                     AVVVPQQGGQLWVPGGPEASPSGPQPYHYNPLWVPLPLALTLWPQVKSLGGPSGPSCHR
                     MVAWKRQLILPGGPHESTRDYIYYNDVYAFNLDTFTWSKLSPSSTGPLPRSGCQMSVI
                     PQQGIVIYGGYSKQRVKKDVDEGTRHSPMPLLKPPPPGEEDEWVWTRMNPSGVKPTIPKS
                     GPSVAMAPNHQTLLPGGVCDEEEEESLAGEPPNDLYFYDATRNPWPPGQLKGPPKSEKK
                     KKKKGKEEEPEEGGSKLACQGAGIQAPVEVVKEVVIPIGIVVIIKQVLTAPGSAGQPQS
                     EDEDSLEEAGSPTPGPCPRSNAMLAVKHGVLYVYGGMPEAGDRQVTLSDLHCLDLHRM
                     EAWKALVEMDPPLTQEWLEEIDSEEDSEEVEGAEGGGEDEDEEGSGEESGAED"
BASE COUNT       440 a    490 c    623 g    358 t
ORIGIN
```

Alignment scores:

Quality:	7.00	Length:	7
Ratio:	1.000	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

Alignment block:

US-09-126-945B-2 COPY_46_54 x AR055250/rev ..

Align seq 1/1 to reverse of: AR055250 from: 1 to: 1911

```
       1 SerProProAlaThrProGlu 7
         |||||||||||||||||||||
    1475 TCTCCTCCAGCCACTCCTGAG 1455
```

```
OM of: US-09-126-945B-2_COPY_46_54 *** N_Geneseq_0601 *    qy  1 - 9  ps

Date: Sep 14, 2001  9:48 PM

About: Results were produced by the GenCore [illegible]
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
  MODEL framer_p2n.model -DEV xlp
  -Q "/usr/local/GSFdata/cgn2/45/gcg_temp_[illegible]/app_query.fasta_1.267
  -DB N_Geneseq_0601 -FORMT fastap -SUFFIX olip2n.rng -GAPOP 4.500
  -GAPEXT 0.050 -MINMATCH 0.100 -LOOPCL 0.000 -LOOPEXT 0.000
  -QGAPOP 4.500 -QGAPEXT 0.050 -XGAPOP 60.000 -XGAPEXT 60.000
  -FGAPOP 6.000 -FGAPEXT 7.000 -YGAPOP 60.000 -YGAPEXT 60.000
  -DELOP 6.000 -DELEXT 7.000 -START 1 -MATRIX olig3
  -TRANS human40.cdi -LIST 45 -DOCALIGN 200 -THR_SCORE quality
  -THR_MIN 1 -ALIGN 15 -MODE LOCAL -OUTFMT pir -NORM ext  MINLEN 0
  MAXLEN 2000000000 -USPR=/usr/local/GSFdata/cgn2/45/gcg_temp_[illegible]/app_query.fasta_1.267
  -LONGLOG -NO_XLPXY -WAIT -THREADS 1

Search information block:
Query: US-09-126-945B-2_COPY_46_54
Query length: 9
Database: N_Geneseq_0601:*
Database sequences: 730101
Database length: 413950809
Search time (sec): 680.150000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000


score_list:
Sequence            Strd Orig    ZScore    EScore Len  ! Documentation  ..
/SIDSL/geodata/geneseq/geneseqp/NA2000.DAT:AAF21828 +     9.00   157.43   0.9841     1087  !
/SIDSL/geodata/geneseq/geneseqp/NA2000.DAT:AAF22091 +     9.00   153.82   1.05       1064  !
/SIDSL/geodata/geneseq/geneseqp/NA2001.DAT:AAC84261 +     9.00   153.82   1.05       1064  !
/SIDSL/geodata/geneseq/geneseqp/NA2000.DAT:AAF22092 +     9.00   152.14   1.05       1052  !
/SIDSL/geodata/geneseq/geneseqp/NA1999.DAT:AAA17752 +     9.00   120.51   6.14       254   !
/SIDSL/geodata/geneseq/geneseqp/NA2000.DAT:AAF22471 +     6.00   116.67   159.94     1002  !
/SIDSL/geodata/geneseq/geneseqp/NA2000.DAT:AAF22484 +     6.00   116.00   170.61     256   !
/SIDSL/geodata/geneseq/geneseqp/NA2001.DAT:AAB72917 +     6.00   114.82   205.41     1056  !
/SIDSL/geodata/geneseq/geneseqp/NA1993.DAT:AAX61258 +     6.00   114.50   208.70     264   !
/SIDSL/geodata/geneseq/geneseqp/NA1999.DAT:AAX11323 +     6.00   114.47   232.13     300   !
/SIDSL/geodata/geneseq/geneseqp/NA2000.DAT:AAA71156 +     6.00   113.67   232.13     296   !
/SIDSL/geodata/geneseq/geneseqp/NA1996.DAT:AAX22538 +     6.00   112.76   261.88     345   !
/SIDSL/geodata/geneseq/geneseqp/NA1998.DAT:AAV41826 +     6.00   112.75   261.88     346   !
/SIDSL/geodata/geneseq/geneseqp/NA1999.DAT:AAZ00708 +     6.00   112.75   261.88     346   !
/SIDSL/geodata/geneseq/geneseqp/NA2000.DAT:AAA74148 +     6.00   112.75   261.88     346   !
/SIDSL/geodata/geneseq/geneseqp/NA2001.DAT:AAF67558 +     6.00   111.92   290.60     393   !
/SIDSL/geodata/geneseq/geneseqp/NA1999.DAT:AAX24502 +     6.00   111.55   304.68     416   !
/SIDSL/geodata/geneseq/geneseqp/NA1999.DAT:AAA14537 +     6.00   111.55   304.68     416   !
/SIDSL/geodata/geneseq/geneseqp/NA1999.DAT:AAX24501 +     6.00   111.55   304.68     416   !
/SIDSL/geodata/geneseq/geneseqp/NA1999.DAT:AAX24629 +     6.00   111.55   304.68     416   !
/SIDSL/geodata/geneseq/geneseqp/NA2000.DAT:AAF78863 +     6.00   111.47   311.69     420   !
/SIDSL/geodata/geneseq/geneseqp/NA2001.DAT:AAF85522 +     6.00   110.83   351.21     465   !
/SIDSL/geodata/geneseq/geneseqp/NA2000.DAT:AAF79211 +     6.00   110.81   351.86     466   !
/SIDSL/geodata/geneseq/geneseqp/NA2000.DAT:AAF15711 +     6.00   110.50   418.55     554   !
/SIDSL/geodata/geneseq/geneseqp/NA2000.DAT:AAA15290 +     6.00   109.98   472.70     624   !
/SIDSL/geodata/geneseq/geneseqp/NA1999.DAT:AAV86412 +     6.00   109.76   493.21     648   !
/SIDSL/geodata/geneseq/geneseqp/NA2000.DAT:AAC47688 +     6.00   108.97   424.00     610   !
/SIDSL/geodata/geneseq/geneseqp/NA2000.DAT:AAF55625 +     6.00   108.85   425.23     621   !
/SIDSL/geodata/geneseq/geneseqp/NA1999.DAT:AAX24502 +     6.00   108.63   443.28     652   !
/SIDSL/geodata/geneseq/geneseqp/NA1998.DAT:AAX61607 +     6.00   107.97   482.04     722   !
/SIDSL/geodata/geneseq/geneseqp/NA2000.DAT:AAA41604 +     6.00   107.88   487.69     729   !
/SIDSL/geodata/geneseq/geneseqp/NA2000.DAT:AAA50544 +     6.00   107.84   522.80     768   !
/SIDSL/geodata/geneseq/geneseqp/NA2000.DAT:AAF29444 +     6.00   107.75   543.35     797   !
/SIDSL/geodata/geneseq/geneseqp/NA1998.DAT:AAX06770 +     6.00   107.68   547.83     813   !
/SIDSL/geodata/geneseq/geneseqp/NA1999.DAT:AAX06774 +     6.00   107.67   548.48     813   !
/SIDSL/geodata/geneseq/geneseqp/NA2001.DAT:AAF71803 +     6.00   107.63   558.89     821   !
/SIDSL/geodata/geneseq/geneseqp/NA2001.DAT:AAF71804 +     6.00   107.63   558.89     821   !
/SIDSL/geodata/geneseq/geneseqp/NA2000.DAT:AAF17715 +     6.00   107.52   566.31     832   !
/SIDSL/geodata/geneseq/geneseqp/NA2000.DAT:AAA59060 +     6.00   107.49   567.39     834   !
/SIDSL/geodata/geneseq/geneseqp/NA2000.DAT:AAX06773 +     6.00   106.1    633.10     934   !
/SIDSL/geodata/geneseq/geneseqp/NA2000.DAT:AAA24491       6.00   106.03   636.27     938   !
/SIDSL/geodata/geneseq/geneseqp/NA2001.DAT:AAF62773 +     6.00   105.9    643.61     949   !
/SIDSL/geodata/geneseq/geneseqp/NA2000.DAT:AAA89465 +     6.00   105.85   646.48     953   !
/SIDSL/geodata/geneseq/geneseqp/NA2000.DAT:AAA49787 +     6.00   105.85   653.14     1002  !
/SIDSL/geodata/geneseq/geneseqp/NA2000.DAT:AAA38858 +     6.00   105.85   653.14     1002  !

Seq1 name: /SIDSL/geodata/geneseq/geneseqp/NA2000.DAT:AAF21828

Seq1 documentation block:
ID   AAF21828 standard; DNA; 1087 BP.
XX
AC   AAF21828;
XX
DT   27-MAR-2001  (first entry)
XX
DE   Human breast and ovarian cancer associated antigen gene SEQ ID 27.
XX
KW   Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
KW   nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW   antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
KW   antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
KW   Addison's disease; allergy; autoimmune haemolytic anaemia;
KW   autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW   multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW   cardiovascular disorder; wound healing; neurological disease; ds.
XX
OS   Homo sapiens.
XX
PN   WO200051617-A1.
XX
PD   31-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05881.
XX
PR   12-MAR-1999;  99US-0124270.
XX
PA   (HUMA ) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Ruben SM;
XX
DR   WPI; [illegible]-58.
DR   P-PSDB; AAB58925.
XX
PT   New human breast and ovarian cancer associated gene sequences and the
PT   polypeptides encoded by these genes, useful in the prevention,
PT   treatment and diagnosis of cancer, immune disorders, cardiovascular
PT   disorders and neurological diseases -
XX
PS   Claim 1; Page 642-643; 1299pp; English.
XX
CC   Sequences AAF21811 - AAF22083 represent DNA sequences encoding human
CC   proteins AAB59711 - AAB59728. The DNA and protein sequences are
CC   associated with breast and ovarian cancer. Included in the invention are
CC   sequences AAF22092 - AAF22484 and AAB59129 which are used in the
CC   isolation and characterisation of the DNA and protein sequences of the
CC   invention. The breast and ovarian cancer associated DNA, protein, agonist
CC   and antagonist sequences exhibit cytostatic, immunosuppressive,
CC   nootropic, neuroprotective, antiviral, antiallergic, hepatotropic,
CC   antidiabetic, antiinflammatory, antiulcer, vulnerary, anticonvulsant,
CC   antibacterial, antifungal, antiparasitic and cardiant activity. The
CC   polynucleotide and protein sequences are used in the diagnosis of cancer,
CC   particularly breast and ovarian cancer. The nucleic acid sequences,
CC   proteins, agonists and agonists may also be used in the diagnosis,
CC   prevention and treatment of immune disorders e.g. Addison's disease,
CC   allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC   diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC   arthritis, ulcerative colitis, cardiovascular disorders such as
CC   myocardial ischaemia, wound healing, neurological diseases such as
CC   cerebral ischaemia and epilepsy, and infectious diseases.
XX
SQ   Sequence 1087 BP; 217 A; 347 C; 349 G; 182 T; 4 other;
```

```
alignment scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2 COPY 46_54 x AAF21828  ..

 Align seq 1/1  to: AAF21828  from: 1  to: 1087

        1 SerProProAlaThrProGluGlnGly 9
          |||||||||||||||||||||||||||
       12 AGTCCACCCGCCACGCCCGAGCAGGGC 38

seq name: /STN1/pubdata/geneseq/geneseqNA20...AI.AAZ50691

seq documentation block:
ID   AAZ50691 standard; cDNA; 1894 BP.
XX
AC   AAZ50691;
XX
DT   23-MAY-2000  (first entry)
XX
DE   Human prostate derived Ets factor cDNA.
XX
KW   Prostate Derived Ets factor; PDEF; human; chromosome 6p21.3; cancer;
KW   loss of heterozygosity; chromosomal translocation; linkage analysis;
KW   cytostatic; cardiant; immunosuppressive; cerebroprotective; fungicide;
KW   antibacterial; vulnerary; neuroprotective; antiparkinsonian; nootropic;
KW   anabolic; antiinflammatory; anorectic; hybridisation probe; forensic;
KW   tumour marker; diagnosis, treatment, prostate cancer, blood coagulation,
KW   autoimmune disorder; haematopoietic; immune/nervous system; stroke;
KW   neoplasm; microbial infection; tissue regeneration; heart attack;
KW   scarring; food additive; preservative; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             416..1423
FT                   /*tag= a
FT                   /product= "Human prostate derived Ets factor"
FT                   /note= "Expressed in prostate epithelium"
XX
PN   WO200006589-A1.
XX
PD   10-FEB-2000.
XX
PF   02-AUG-1999;  99WO-US17470.
XX
PR   31-JUL-1998;  98US-0126945.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI   Libermann TA; Oettgen JP; Kunsch CA; Endress GA; Rosen CA;
XX
DR   WPI; 2000-195255/17.
DR   P-PSDB; AAY45003.
XX
PT   Novel prostate derived polypeptide, polynucleotide useful for
PT   diagnosis, prevention and treatment of prostate cancer, autoimmune
PT   disorders, microbial infections and also as food additive or
PT   preservative
XX
PS   Claim 4; Fig 1; 132pp; English.
XX
CC   The present cDNA sequence encodes a human prostate derived Ets factor
CC   (PDEF), isolated from human prostate carcinoma cDNA library
CC   (ATCC No.203072). PDEF gene expression is highest in tissues involved in
CC   androgen sensitivity, with enriched expression in prostate epithelium
CC   cells. PDEF gene is mapped to the human chromosome 6p21.3
CC   region that is associated with loss of heterozygosity and chromosomal
CC   translocations in various human cancers. PDEF has cytostatic, cardiant,
CC   immunosuppressive, cerebroprotective, fungicide, antibacterial,
CC   vulnerary, neuroprotective, antiparkinsonian, nootropic, anabolic,
CC   antiinflammatory and anorectic activity. PDEF polynucleotides are useful
CC   in linkage analysis as markers, as hybridisation probes for differential
CC   identification of the tissues or cell types and as polymorphic markers
CC   for forensic purposes. PDEF is useful as prostate specific tumour marker
CC   for the diagnosis and treatment of prostate cancer. PDEF sequences are
CC   useful for treating autoimmune disorders, haematopoietic, blood
CC   coagulation, immune and nervous system disorders, hyperproliferative
CC   disorders like, neoplasms and microbial infections, heart attacks,
CC   stroke, scarring and for tissue regeneration. They are also useful as
CC   food additives or preservatives.
XX
SQ   Sequence 1894 BP; 368 A; 653 C; 571 G; 302 T; 0 other;


alignment scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_46_54 x AAZ50691  ..

 Align seq 1/1  to: AAZ50691  from: 1  to: 1894

        1 SerProProAlaThrProGluGlnGly 9
          |||||||||||||||||||||||||||
      551 AGTCCACCCGCCACGCCCGAGCAGGGC 577

seq name: /STN1/pubdata/geneseq/geneseqNA2001.DAT:AAC83261

seq documentation block:
ID   AAC83261 standard; DNA; 1894 BP.
XX
AC   AAC83261;
XX
DT   16-MAR-2001  (first entry)
XX
DE   Gland specific Ets transcription factor (GSEF) nucleotide sequence.
XX
KW   Transcription factor; gland specific Ets transcription factor; GSEF;
KW   metastatic potential; cancer; tumour; metastasis; breast; prostate;
KW   leukaemia; lymphoma; sarcoma; melanoma; chromosome 6p21.1 6p21.3; ds.
XX
OS   Homo sapiens.
XX
PN   WO200070092-A1.
XX
PD   23-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US13173.
XX
PR   14-MAY-1999;  99US-0134112.
XX
PA   (CHIR-) CHIRON CORP.
XX
PI   Kauffmann F; Xin H; Barrowe G;
XX
DR   WPI; 2001-011019/05.
DR   P-PSDB; AAB49628.
XX
PT   Detecting metastatic and potential metastatic cancerous cells, useful
PT   for diagnosing, prognosing, grading and staging of cancers by detecting
PT   gland specific Ets transcription factor gene product in a biological
PT   sample from a cell
XX
PS   Disclosure; Page 81-83; 95pp; English.
XX
CC   This invention relates to a method for the detection and determination of
CC   the metastatic potential of a cell. The method comprises detecting a
CC   gland specific Ets transcription factor (GSEF) gene product in a test
```

```
CC   sample. Detection of a GSEF gene product in the test sample in amount
CC   lower than that in a normal cell, is indicative of a cell with high
CC   metastatic potential. The method is useful for determining the metastatic
CC   potential of a cell, for the diagnosis and prognosis of cancer as well as
CC   grading and staging of cancers by detecting GSEF expression in a
CC   biological test sample. The method may also be used to monitor patients
CC   having a predisposition to develop a particular cancer. GSEF polypeptides
CC   are useful for producing antibodies, in cancer diagnosis, prognosis,
CC   grading, staging and management of breast and prostate tumours, and in
CC   detecting polymorphisms in the sequence. GSEF genes and proteins are also
CC   useful in gene therapy. GSEF gene product expression levels can be used
CC   in conjunction with any tissue in which an alteration in GSEF gene
CC   product expression levels is associated with development of a
CC   cancer-associated phenotype. Cancers, which can be monitored include
CC   cancers of the prostate, cervix, lung and colon, melanomas, colorectal
CC   adenocarcinomas, Wilm's tumour, retinoblastoma, sarcomas, myosarcomas,
CC   lung carcinomas, leukaemia, and lymphomas. The GSEF gene is located on
CC   human chromosome 6, specifically at 6p21.1-6p21.3. The present sequence
CC   represents the DNA encoding GSEF.
XX
SQ   Sequence 1894 BP; 368 A; 653 C; 571 G; 302 T; 0 other;


alignment_scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_46_54 x AAC83261   ..

 Align seq 1/1  to: AAC83261  from: 1  to: 1864

       1 SerProProAlaThrProGluGlnGly 9
         |||||||||||||||||||||||||||
     551 AGTCCCACCCGCCACGCCCGAGCAGGGC 577

Seq_name: /SIDS2/seqdir/geneseq/geneseqp/NA* : *: AAX40792

seq_documentation_block:
ID   AAC83266 standard; DNA; 3317 BP.
XX
AC   AAC83266;
XX
DT   16-MAR-2001  (first entry)
XX
DE   Gland-specific Ets transcription factor (GSEF) cDNA sequence.
XX
KW   transcription factor; gland-specific Ets transcription factor; GSEF;
KW   metastatic potential; cancer; tumour; metastasis; breast; prostate;
KW   leukaemia; lymphoma; sarcoma; melanoma; chromosome 6p21.1-6p21.3; ss.
XX
OS   Homo sapiens.
XX
PN   WO200070092-A1.
XX
PD   23-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US13173.
XX
PR   14-MAY-1999;   99US-0134112.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Kaufmann J,  Xin H,  Harrowe G;
XX
DR   WPI; 2001-041019/05.
DR   P-PSDB; AAB49633.
XX
PT   Detecting metastatic and potential metastatic cancerous cells, useful
PT   for diagnosing, prognosing, grading and staging of cancers by detecting
PT   gland-specific Ets transcription factor gene product in a biological
PT   sample from a cell -
XX
PS   Disclosure; Fig 2; 95pp; English.
XX
CC   This invention relates to a method for the detection and determination of
CC   the metastatic potential of a cell. The method comprises detecting a
CC   gland specific Ets transcription factor (GSEF) gene product in a test
CC   sample. Detection of a GSEF gene product in the test sample in amount
CC   lower than that in a normal cell, is indicative of a cell with high
CC   metastatic potential. The method is useful for determining the metastatic
CC   potential of a cell, for the diagnosis and prognosis of cancer as well as
CC   grading and staging of cancers by detecting GSEF expression in a
CC   biological test sample. The method may also be used to monitor patients
CC   having a predisposition to develop a particular cancer. GSEF polypeptides
CC   are useful for producing antibodies, in cancer diagnosis, prognosis,
CC   grading, staging and management of breast and prostate tumours, and in
CC   detecting polymorphisms in the sequence. GSEF genes and proteins are also
CC   useful in gene therapy. GSEF gene product expression levels can be used
CC   in conjunction with any tissue in which an alteration in GSEF gene
CC   product expression levels is associated with development of a
CC   cancer-associated phenotype. Cancers, which can be monitored include
CC   cancers of the prostate, cervix, lung and colon, melanomas, colorectal
CC   adenocarcinomas, Wilm's tumour, retinoblastoma, sarcomas, myosarcomas,
CC   lung carcinomas, leukaemia, and lymphomas. The GSEF gene is located on
CC   human chromosome 6, specifically at 6p21.1-6p21.3. The present sequence
CC   represents the cDNA encoding GSEF.
XX
SQ   Sequence 3317 BP; 710 A; 1026 C; 970 G; 611 T; 0 other;


alignment_scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_46_54 x AAC83266   ..

 Align seq 1/1  to: AAC83266  from: 1  to: 3317

       1 SerProProAlaThrProGluGlnGly 9
         |||| |||||||||||||||||||||
    1950 AGTCCACCCGCCACGCCCGAGCAGGGC 1976

Seq_name: /SIDS2/seqdir/geneseq/geneseqp/NA* : *: AAX40792

seq_documentation_block:
ID   AAX40792 standard; DNA; 26 BP.
XX
AC   AAX40792;
XX
DT   16-JUN-1999  (first entry)
XX
DE   oligonucleotide used for screening DCMIase binding discrimination.
XX
KW   DNA cytosine methyltransferase; DCMIase; binding discrimination;
KW   5'-S-methylcytosine, allosteric site, synthetic inhibitor; cancer cell;
KW   proliferation; development disorder; Huntington's disease;
KW   Down's syndrome; Hox gene; ss.
XX
OS   Synthetic.
XX
PN   WO9912027-A1.
XX
PD   11-MAR-1999.
XX
PF   12-JUN-1998;   98WO-US12451.
XX
PR   29-AUG-1997;   97US-0057411.
XX
PA   (REGE ) UNIV CALIFORNIA.
XX
PI   Flynn J,  Reich N;
XX
```

DR WPI; 1999-205256/17.
XX
PT DNA cytosine methyltransferase modulator containing 5'-C
PT methylcytosine used to inhibit methylation of DNA, and
PT proliferation of cancer cells
XX
PS Example 1; Fig 7; 114pp; English.
XX
CC oligonucleotides AAX40733-815 represent cloned and sequenced isolates
CC from pooled generations used for screening DNA cytosine
CC methyltransferase (DCMTase) binding discrimination. The oligonucleotides
CC are double stranded, and the guanine containing strand is shown. The
CC specification describes a synthetic oligonucleotide comprising a C'5
CC methylcytosine which recognizes and binds an allosteric site on DCMTase,
CC thereby modulating DCMTase activity associated with the allosteric site.
CC The synthetic inhibitor can be used to inhibit methylation of DNA. It
CC can also be used to inhibit proliferation of cancer cells. The inhibitor
CC can also be used to treat a disorder of development, which is linked to
CC a genetic locus regulated by methylation, such as Huntington's disease,
CC Down's syndrome, and disorders associated with a Box gene.
XX
SQ Sequence 26 BP; 3 A; 1 C; 17 G; 5 T; 0 other;

alignment scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2_COPY_46_54 x AAX40792/rev ..

Align seq 1/1 to reverse of: AAX40792 from: 1 to: 26

1 SerProProAlaThrPro 6
|||||||||||||||||||
22 TCACCACCCGCCACTCCC 5

Seq name: /cgi6/ptodata/genseq/genseq_DNA_new/AAX40792.rng

Seq documentation block:
ID AAC27474 standard; cDNA; 189 BP.
XX
AC AAC27474;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 41549.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST-) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500481/45.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 41549; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 40 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
SQ Sequence 189 BP; 21 A; 50 C; 66 G; 46 T; 6 other;

alignment scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2_COPY_46_54 x AAC27474/rev ..

Align seq 1/1 to reverse of: AAC27474 from: 1 to: 189

2 ProProAlaThrProGlu 7
|||||||||||||||||||
98 CCACCCGCCACTCCGGAA 81

Seq name: /cgi6/ptodata/genseq/genseq_DNA2000.DAT:AAC22981

Seq documentation block:
ID AAC22981 standard; cDNA; 216 BP.
XX
AC AAC22981;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 27056.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST-) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500481/45.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 27056; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 40 different tissues. EST

CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
SQ Sequence 216 BP; 56 A; 60 C; 71 G; 29 T; 0 other;

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_46_54 x AAC22981 ..

Align seq 1/1 to: AAC22981 from: 1 to: 216

1 SerProProAlaThrPro 6
|||||||||||||||||||
81 TCTCCGCCAGCCACTCCA 98

seq_name: [illegible]

seq_documentation_block:
ID AAF72017 standard; DNA; 259 BP.
XX
AC AAF72017;
XX
DT 30-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:529.
XX
KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;
KW fine chemical production; microorganism; organic acid; nucleoside;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;
KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme; ds.
XX
OS Corynebacterium glutamicum.
XX
PN WO200103843-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-IB00923.
XX
PR 25-JUN-1999; 99US-0141031.
PR 01-JUL-1999; 99DE-1030476.
PR 02-JUL-1999; 99US-0142181.
PR 08-JUL-1999; 99DE-1031415.
PR 08-JUL-1999; 99DE-1031418.
PR 08-JUL-1999; 99DE-1031419.
PR 08-JUL-1999; 99DE-1031420.
PR 08-JUL-1999; 99DE-1031424.
PR 08-JUL-1999; 99DE-1031428.
PR 08-JUL-1999; 99DE-1031434.
PR 08-JUL-1999; 99DE-1031435.
PR 08-JUL-1999; 99DE-1031443.
PR 08-JUL-1999; 99DE-1031453.
PR 08-JUL-1999; 99DE-1031457.
PR 08-JUL-1999; 99DE-1031465.
PR 08-JUL-1999; 99DE-1031478.
PR 08-JUL-1999; 99DE-1031510.
PR 08-JUL-1999; 99DE-1031541.
PR 08-JUL-1999; 99DE-1031573
PR 08-JUL-1999; 99DE-1031592.
PR 08-JUL-1999; 99DE-1031632.
PR 08-JUL-1999; 99DE-1031644.
PR 08-JUL-1999; 99DE-1031636.
PR 09-JUL-1999; 99DE-1032125.
PR 09-JUL-1999; 99DE-1032126.
PR 09-JUL-1999; 99DE-1032130.
PR 09-JUL-1999; 99DE-1032186.
PR 09-JUL-1999; 99DE-1032206.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032228.
PR 09-JUL-1999; 99DE-1032229.
PR 09-JUL-1999; 99DE-1032230.
PR 14-JUL-1999; 99DE-1032922.
PR 14-JUL-1999; 99DE-1032926.
PR 14-JUL-1999; 99DE-1032928.
PR 14-JUL-1999; 99DE-1033004.
PR 14-JUL-1999; 99DE-1033005.
PR 14-JUL-1999; 99DE-1033006.
PR 12-AUG-1999; 99US-0148618.
PR 27-AUG-1999; 99DE-1040764.
PR 27-AUG-1999; 99DE-1040765.
PR 27-AUG-1999; 99DE-1040766.
PR 27-AUG-1999; 99DE-1040832.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041380.
PR 31-AUG-1999; 99DE-1041394.
PR 31-AUG-1999; 99DE-1041396.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042077.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042124.
PR 03-SEP-1999; 99DE-1042129.
PR 09-MAR-2000; 2000US-0187970.
XX
PA (BADI) BASF AG.
XX
PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G,
XX
DR WPI; 2001-137957/14.
DR P-PSDB; AAB79898.
XX
PT Nucleic acids from Corynebacterium glutamicum encoding metabolic
PT pathway proteins, useful for producing fine chemicals in
PT microorganisms, including organic acids, nonproteinogenic amino acids,
PT and purine and pyrimidine bases -
XX
PS Claim 4; Page 920; 1737pp; English.
XX
CC AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
CC pathway (MP) proteins given in AAB79604 to AAB80211. The C. glutamicum
CC MP nucleic acids are useful for the production of fine chemicals
CC in microorganisms, including organic acids, nonproteinogenic amino
CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
CC compounds, vitamins, cofactors, polyketides and enzymes.
XX
SQ Sequence 259 BP; 64 A; 77 C; 65 G; 54 T; 0 other;

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_46_54 x AAF72017/rev ..

Align seq 1/1 to reverse of: AAF72017 from: 1 to: 259

```
        2 ProProAlaThrProGln 7
          ||||||||||||||||||
      182 CCTCCAGCAACGGCCTGAC 165

Seq name: [illegible]

Seq documentation block:
ID   AAQ61258 standard; DNA; 264 BP.
XX
AC   AAQ61258;
XX
DT   16-MAR-1994  (first entry)
XX
DE   Human brain Expressed Sequence Tag EST01744.
XX
KW   Gene transcription product; genetic markers; tagging; in vivo;
KW   transcription; mapping; locations; chromosomes; chromosomal; ss.
XX
OS   Homo sapiens.
XX
PN   WO9416178 A.
XX
PD   19-AUG-1994.
XX
PF   12-FEB-1993;   94WO-US01294.
XX
PR   12-FEB-1992;   92US-0834795.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
PI   Adams MD,  Moreno RF,  Venter CJ;
XX
DR   WPI; 1994-272882/44.
XX
PT   Enriched oligonucleotides and corresp. sequences  used as
PT   markers for human genes transcribed in vivo, facilitate tagging
PT   of most human genes
XX
PS   Example 4; Page 463; 500pp; English.
XX
CC   The Expressed Sequence Tag was isolated from a human brain cDNA
CC   library as part of a large set of ESTs which can be used as markers
CC   for human genes transcribed in vivo. They can be used to facilitate
CC   tagging of most human genes, for mapping locations of expressed genes
CC   on chromosomes, for individual or forensic identification, for mapping
CC   locations of disease associated genes, for identification of tissue
CC   type, and for prepn. of antisense sequences, probes and constructs.
CC   EST01744 has been putatively identified as NAD(P)-transhydrogenase
CC   (B-specific). See also AAQ59041 Q61440.
XX
SQ   Sequence 264 BP; 48 A; 68 C; 77 G; 69 T; 2 other;


alignment scores:
             Quality:     6.00              Length:       6
               Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2 COPY_46_54 x AAQ61258/rc?     .

 Align seq 1/1  to reverse of: AAQ61258  from: 1  to: 264

        1 SerProProAlaThrPro 6
          ||||||||||||||||||
      186 AGTCCTCCAGCAACCCCA 169

Seq name: 2 cDNA/Sequence.geneseq-genome-pg:NAT:???:AI:AAZ11323

Seq documentation block:
ID   AAZ11323 standard; DNA; 300 BP.
XX
```

```
AC   AAZ11323;
XX
DT   25-OCT-1999  (first entry)
XX
DE   M. vaccae 65 kd heat shock protein partial DNA sequence.
XX
KW   Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW   dendritic cell maturation; infectious disease; immune disorder; cancer;
KW   respiratory system; mycobacterial infection; allergy; tuberculosis;
KW   leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW   dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW   squamous cell carcinoma; melanoma; ss.
XX
OS   Mycobacterium vaccae.
XX
PN   WO9932634-A2.
XX
PD   01-JUL-1999.
XX
PF   23-DEC-1998;   98WO-NZ00189.
XX
PR   04-DEC-1998;   98US-0205426.
PR   23-DEC-1997;   97US-0996624.
PR   23-DEC-1997;   97US-0997080.
PR   23-DEC-1997;   97US-0997362.
PR   11-JUN-1998;   98US-0095855.
PR   17-SEP-1998;   98US-0156181.
XX
PA   (GENE ) GENESIS RES & DEV CORP LTD.
XX
PI   Prestidge RL,  Skinner MA,  Tan P,  Visser ES,  Watson JD;
XX
DR   WPI; 1999-430163/36.
DR   P-PSDB; AAY14622.
XX
PT   Enhancing immune response to an antigen
XX
PS   Example 14; Page 176; 243pp; English.
XX
CC   The invention provides heat-killed Mycobacterium vaccae, or recombinant
CC   M. vaccae proteins. The M. vaccae proteins may be employed to activate
CC   T cells and natural killer cells, to stimulate the production of
CC   cytokines, to enhance the expression of co-stimulatory molecules on
CC   dendritic cells and monocytes, and to enhance dendritic cell maturation
CC   and function. The proteins can be expressed by standard recombinant
CC   methodology. Pharmaceutical compositions comprising the proteins or
CC   nucleic acid sequences encoding the proteins can be used for the
CC   treatment, prevention, and detection of disorders including infections
CC   diseases, immune disorders and cancer. In particular, the compounds and
CC   methods are used for treatment of diseases of the respiratory system,
CC   such as mycobacterial infections, asthma, allergies, tuberculosis,
CC   leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC   psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC   alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC   carcinoma and melanoma.
XX
SQ   Sequence 300 BP; 60 A; 104 C; 98 G; 48 T; 0 other;


alignment scores:
             Quality:     6.00              Length:       6
               Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2 COPY_46_54 x AAZ11323   ..

 Align seq 1/1  to: AAZ11323  from: 1  to: 300

        1 SerProProAlaThrPro 6
          ||||||||||||||||||
      257 TCTCCCCCGGGCACACCC 274
```

```
Seq_name: C:\Cirn\Tir\Seq\Seq_Inbox\pCA...\US-09-126A-01156

Seq_documentation_block:
ID   AAA01156 standard; cDNA; 300 BP.
XX
AC   AAA01156;
XX
DT   19-MAY-2000 (first entry)
XX
DE   Human colon cancer cell line polynucleotide sequence SEQ ID NO:1147.
XX
KW   Human; colon cancer; tumour; diagnosis; gene expression product;
KW   probe; detection; cancerous state; metastasis; identification;
KW   breast cancer; oestrogen receptor-positive breast cancer; therapy;
KW   oestrogen receptor-negative breast cancer; lung cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO99/58675 A2.
XX
PD   18-NOV-1999.
XX
PF   13-MAY-1999;   99WO-US10602.
XX
PR   14-MAY-1998;   98US-0085426.
PR   15-MAY-1998;   98US-0085547.
PR   15-MAY-1998;   98US-0085696.
PR   21-OCT-1998;   98US-0105234.
PR   27-OCT-1998;   98US-0105877.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Williams LT, Escobedo J, Innis MA, Garcia PD, Suddath Klinger J;
PI   Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI   Lawson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Lahat I;
PI   Leshkowitz D, Kita D, Garcia V, Jones LW, Stuche-Crain B;
XX
DR   WPI; 2000-126369/11.
XX
PT   Polynucleotide library used to determine cancerous states of mammalian
PT   cells
XX
PS   Claim 1; Page 456; 1097pp; English.
XX
CC   AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
CC   libraries constructed from human colon cancer cell lines. The present
CC   invention also describes a method of detecting differentially expressed
CC   genes correlated with a cancerous state of a mammalian cell, comprising
CC   detecting at least one differentially expressed gene product in a test
CC   sample derived from a cell suspected of being cancerous, where detection
CC   of the differentially expressed gene product is correlated with a
CC   cancerous state of the cell from which the test sample was derived.
CC   The polynucleotides sequences can be used in a method for detecting
CC   differentially expressed genes correlated with a cancerous state of a
CC   mammalian cell. The polynucleotides can also be used as probes for
CC   detecting and mapping related genes. They can be used in diagnosis and
CC   prognosis of diseases and disorders (e.g. identification of
CC   pre-metastatic or metastatic cancerous states, stages of cancer, or
CC   responsiveness of cancer to therapy). This is particularly for breast
CC   cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
CC   negative breast cancer, lung cancer, and colon cancer.
XX
SQ   Sequence 300 BP; 95 A; 114 C; 33 G; 58 T; 0 other;


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_46_54 x AAA01156    ..

 Align seq 1/1  to: AAA01156  from: 1  to: 300

        1 SerProProAlaThrPro 6
          || ||||||||||||||||
      112 TCTCCAGCTCCAACCCCA 129

Seq_name: C:\Cirn\Tir\Seq\Seq_Inbox\pCA...\US-09-126A-AAX22530

Seq_documentation_block:
ID   AAX22530 standard; DNA; 345 BP.
XX
AC   AAX22530;
XX
DT   21-MAY-1999 (first entry)
XX
DE   Murine GDF 15 DNA
XX
KW   GDF-15; mouse; growth differentiation factor; detection; ovarian; colon;
KW   treatment; cell proliferative disorder; immunological disorder; disease;
KW   muscle; musculodegenerative disorder; tissue repair; burn; contraceptive;
KW   premature labour; fetal growth; cancer; gonadal; adrenal; tumour;
KW   endometrial; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/qualifiers
FT   CDS             1..345
FT                   /*tag= a
FT                   /product= "GDF-15"
XX
PN   WO9906445 A1.
XX
PD   11-FEB-1999.
XX
PF   24-JUL-1998;   98WO-US15455.
XX
PR   31-JUL-1997;   97US-0054060.
XX
PA   (JOHJ ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX
PI   Hsiao E, Huynh TV, Lee S, Rankin C, Sebald S;
XX
DR   WPI; 1999-154705/14.
DR   P-PSDB; AAW93161.
XX
PT   A new growth differentiation factor, factor-15 (GDF 15) - useful to
PT   detect and treat cell proliferative and immunological disorders
XX
PS   Claim 3; Fig 1; 38pp; English.
XX
CC   This sequence encodes a novel mouse growth differentiation factor,
CC   GDF-15. The products described in the invention may be used to detect
CC   or treat cell proliferative or immunological disorders, including disease
CC   processes involving muscle such as musculodegenerative disorders or
CC   tissue repair following trauma or burns, in contraceptive regimens, to
CC   prevent premature labour, to enhance fetal growth or development, or to
CC   treat various cancers including gonadal and adrenal tumours, endometrial
CC   and ovarian tumours, and colon cancer.
XX
SQ   Sequence 345 BP; 58 A; 111 C; 109 G; 67 T; 0 other;


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_46_54 x AAX22530    ..

 Align seq 1/1  to: AAX22530  from: 1  to: 345
```

```
        1 SerProProAlaThrPro 6
          ||||||||||||||||||
      245 TCCCCTCCAGCTACACCC 262

Seq name: gi|1185|gb|AAV31825.1|AAV31825

seq documentation block:
ID   AAV31825 standard; cDNA; 446 BP.
XX
AC   AAV31825;
XX
DT   09-SEP-1998 (first entry)
XX
DE   Mutant Aspergillus oryzae 1204.3.3 rescued locus.
XX
KW   Rescued locus; mutant; metabolite; ss.
XX
OS   Aspergillus oryzae.
XX
PN   WO9811203-A1.
XX
PD   19-MAR-1998.
XX
PF   12-SEP-1997;   97WO-US16604.
XX
PR   13-SEP-1996;   96US-0713312.
XX
PA   (NOVO ) NOVO NORDISK BIOTECH INC.
XX
PI   Brody H, Hansen K, Lamsa MH, Yaver DS;
XX
DR   WPI; 1998-207372/18.
XX
PT   Polypeptide production - using a mutant cell with a nucleic acid
PT   construct introduced into a locus
XX
PS   claim 48; Fig 20; 195pp; English.
XX
CC   This is the nucleotide sequence of a rescued locus from the mutant
CC   Aspergillus oryzae 1204.3.3, used in the method of the invention
CC   which involves the use of a mutant cell with a nucleic acid
CC   construct introduced into a locus. The methods can be used for
CC   producing metabolites.
XX
SQ   Sequence 446 BP; 95 A; 66 C; 70 G; 115 T; 0 other;


alignment scores:
             Quality:     6.00              Length:       6
               Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2 (COPY_46_54 x AAV31825/rev    ..

 Align seq 1/1  to reverse of: AAV31825  from: 1  to: 446

        1 SerProProAlaThrPro 6
          ||||||||||||||||||
      150 TCACCGCCTGCAACACCA 133

Seq name: gi|1185|gb|AAV31825.1|AAV31825

seq documentation block:
ID   AAZ20708 standard; cDNA; 446 BP.
XX
AC   AAZ20708;
XX
DT   03-DEC-1999 (first entry)
XX
DE   A. oryzae 1204.3.3 locus DNA sequence.
XX
KW   Protein production; mutant protein; DNA locus; ss
```

```
XX
OS   Aspergillus oryzae.
XX
PN   US5958727-A.
XX
PD   28-SEP-1999.
XX
PF   12-SEP-1997;   97US-0928692.
XX
PR   13-SEP-1996;   96US-0713312.
XX
PA   (NOVO ) NOVO NORDISK BIOTECH INC.
XX
PI   Brody H, Lamsa M, Hansen K, Yaver DS;
XX
DR   WPI; 1999-561041/47.
XX
PT   Modifying the production of a polypeptide
XX
PS   Claim 48; Fig 20; 129pp; English.
XX
CC   This sequence represents an Aspergillus oryzae DNA locus. The invention
CC   relates to a method of producing a polypeptide, comprising:
CC   (a) cultivating a mutant cell under conditions conducive for production
CC   of the polypeptide, where: (i) the mutant cell is related to a parent
CC   cell, which comprises a first DNA sequence encoding the polypeptide, by
CC   the introduction of a nucleic acid construct into the genome of the
CC   parent cell at a locus which is not within the first DNA sequence, and
CC   not within a second DNA which encodes a protein that negatively regulates
CC   transcription, translation or secretion of the polypeptide, and not
CC   within a third DNA sequence encoding a protease which hydrolyses the
CC   polypeptide under the conditions; and (ii) the mutant cell produces more
CC   of the polypeptide than the parent cell when both cells are cultivated
CC   under the conditions; and (b) recovering the polypeptide. The method is
CC   useful for producing polypeptides encoded by a DNA sequence which does
CC   not contain the mutation.
XX
SQ   Sequence 446 BP; 95 A; 66 C; 70 G; 115 T; 0 other;


alignment scores:
             Quality:     6.00              Length:       6
               Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2 (COPY_46_54 x AAZ20708/rev    ..

 Align seq 1/1  to reverse of: AAZ20708  from: 1  to: 446

        1 SerProProAlaThrPro 6
          ||||||||||||||||||
      150 TCACCGCCTGCAACACCA 133

Seq name: gi|2000|gb|AAC04246.1|AAC04246

seq documentation block:
ID   AAC04246 standard; cDNA; 477 BP.
XX
AC   AAC04246;
XX
DT   06-SEP-2000 (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 8321.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033410-A2.
XX
PD   06-SEP-2000.
```

XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GESI) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano I;
XX
DR WPI; 2000-506381/45.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures
XX
PS Claim 1; SEQ ID 8321; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
SQ Sequence 377 BP; 69 A; 150 C; 95 G; 60 T; 3 other;

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_46_54 x AAC04246

Align seq 1/1 to: AAC04246 from: 1 to: 377

1 SerProProAlaThrPro 6
|||||||||||||||||
249 AGTCCTCCGGCGACCCCG 266

[illegible]

Date: Sep 14, 2001 9:53 PM

[illegible]
Copyright (c) [illegible]

Command line parameters:
-MODEL trans [illegible]
[illegible]
-DB Issued_Patents_NA -QFMT fastap -SUFFIX olip2n.rni
-GAPOP 4.500 -GAPEXT=0.050 -MINMATCH=0.100 -[illegible]
-LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP 60.000
-XGAPEXT=60.000 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000
-YGAPEXT 60.000 -DELOP=6.000 -DELEXT=7.000 -SIART=1
-MATRIX oligo -TRANS human40.cdi -LIST=45 -D=ALIGN 200
-THR_SCORE quality -THR_MIN=1 -ALIGN 15 -MODE LOCAL -OUTFMT pts
-NORM=xt -MINLEN=0 -MAXLEN 200000000
-USER=US09126945_*CGN1_1_91 -NCPU=6 -[illegible]
-WAIT -THREADS=1

Search information block:
Query: US 09-126-945B-2_COPY_46_54
Query length: 9
Database: Issued_Patents_NA:*
Database sequences: 324599
Database length: 94655562
Search time (sec): 322.720000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000

score_list:
Sequence Strd Orig Score Ident Length Query_match_len ...

[illegible]

[illegible]

seq_documentation_block:
; Sequence 861, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Rebischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; [illegible]
; [illegible]
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 31-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: [illegible]
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: [illegible]
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 861:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 611 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; [illegible]
; ORIGINAL SOURCE:
; ORGANISM: [illegible]
US 08 998 416 861

[illegible]
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
[illegible]

Align [illegible] US-08-998-416-861 from: 1 to: 614

[illegible] AlaGluProGlu 7

```
          ||||||||||||||||||||
      143 TCGCCGCCCCCCCACCCCCGAG 123

seq name: /opz 6/prodata/2/inq/96 COMB seq:US 08 890 980 26

seq documentation block:
; Sequence 26, Application US/08890980
; Patent No. 5998141
;  GENERAL INFORMATION:
;     APPLICANT:  Acton, Susan L.
;     TITLE OF INVENTION:  SR BI NUCLEIC ACIDS AND USES THEREFOR
;     NUMBER OF SEQUENCES:  86
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  FOLEY, HOAG & ELIOT LLP
;       STREET:  One Post Office Square
;       CITY:  Boston
;       STATE:  MA
;       COUNTRY:  USA
;       ZIP:  02109-2170
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/890,980
;       FILING DATE:  10-JUL-1997
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Arnold, Beth E.
;       REGISTRATION NUMBER:  35,430
;       REFERENCE/DOCKET NUMBER:  MIA 005.01
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617-832-1000
;       TELEFAX:  617-832-7000
;   INFORMATION FOR SEQ ID NO:  26:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  160 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-890-980-26


alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_46_54 x US-08-890-980-26/rev   ..

 Align seq 1/1  to reverse of: US 08-890 980 26  from: 1  to: 160

        3 ProAlaThrProGluGln 8
          ||||||||||||||||||
       80 CCTGCCACTCCCGAGCAG 63

seq name: /opz 6/prodata/2/inq/96 COMB seq:US 08 890 980 66

seq documentation block:
; Sequence 66, Application US/08890980
; Patent No. 5998141
;  GENERAL INFORMATION:
;     APPLICANT:  Acton, Susan L.
;     TITLE OF INVENTION:  SR BI NUCLEIC ACIDS AND USES THEREFOR
;     NUMBER OF SEQUENCES:  86
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  FOLEY, HOAG & ELIOT LLP
;       STREET:  One Post Office Square
;       CITY:  Boston
•       STATE:  MA
        •
;       COUNTRY:  USA
;       ZIP:  02109-2170
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/890,980
;       FILING DATE:  10-JUL-1997
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Arnold, Beth E.
;       REGISTRATION NUMBER:  35,430
;       REFERENCE/DOCKET NUMBER:  MIA 005.01
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617-832-1000
;       TELEFAX:  617-832-7000
;   INFORMATION FOR SEQ ID NO:  66:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  160 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-890-980-66


alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_46_54 x US-08-890-980-66/rev   ..

 Align seq 1/1  to reverse of: US 08-890-980-66  from: 1  to: 160

        3 ProAlaThrProGluGln 8
          ||||||||||||||||||
       80 CCTGCCACTCCCGAGCAG 63

seq name: /opz 6/prodata/2/inq/6A COMB,seq:US 08 890 979 26

seq documentation block:
; Sequence 26, Application US/08890979
; Patent No. 6030778
;  GENERAL INFORMATION:
;     APPLICANT:  Acton, Susan L.
;     APPLICANT:  Ordovas, Jose M.
;     TITLE OF INVENTION:  DIAGNOSTIC ASSAYS AND KITS FOR BODY MASS
;     TITLE OF INVENTION:  DISORDERS
;     NUMBER OF SEQUENCES:  75
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  FOLEY, HOAG & ELIOT LLP
;       STREET:  One Post Office Square
;       CITY:  Boston
;       STATE:  MA
;       COUNTRY:  USA
;       ZIP:  02109-2170
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/893,979
;       FILING DATE:  10-JUL-1997
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Arnold, Beth E.
;       REGISTRATION NUMBER:  35,430
;       REFERENCE/DOCKET NUMBER:  MIA 005.02
```

```
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617-832-1000
;       TELEFAX:  617-832-7000
;   INFORMATION FOR SEQ ID NO:  26:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  160 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-890-979-26


alignment_scores:
             Quality:    6.00              Length:       6
               Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_46_54 x US-08-890-979-26/rev   ..

 Align Seq 1/1  to reverse of: US-08-890-979-26  from: 1  to: 160

        3 ProAlaThrProGluGln 8
          ||||||||||||||||||
       80 CCTGCCACTCCCGAGCAG 63

Seq_name: /cgn_2_ptodata/2_fmn_6A_COMB/seq_US-09-032-894-26

Seq_documentation_block:
; Sequence 65, Application US/08890979
; Patent No. 6030778
;   GENERAL INFORMATION:
;     APPLICANT:  Acton, Susan L.
;     APPLICANT:  Ordovas, Jose M.
;     TITLE OF INVENTION:  DIAGNOSTIC ASSAYS AND KITS FOR BODY MASS
;     TITLE OF INVENTION:  DISORDERS
;     NUMBER OF SEQUENCES:  75
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  FOLEY, HOAG & ELIOT LLP
;       STREET:  One Post Office Square
;       CITY:  Boston
;       STATE:  MA
;       COUNTRY:  USA
;       ZIP:  02109-2170
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/890,979
;       FILING DATE:  10-JUL-1997
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Arnold, Beth E.
;       REGISTRATION NUMBER:  35,430
;       REFERENCE/DOCKET NUMBER:  MIA-005/03
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617-832-1000
;       TELEFAX:  617-832-7000
;   INFORMATION FOR SEQ ID NO:  65:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  160 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-890-979-65


alignment_scores:
             Quality:    6.00              Length:       6
               Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_46_54 x US-08-890-979-65/rev   ..

 Align Seq 1/1  to reverse of: US-08-890-979-65  from: 1  to: 160

        3 ProAlaThrProGluGln 8
          ||||||||||||||||||
       80 CCTGCCACTCCCGAGCAG 63

Seq_name: /cgn_2_ptodata/2_fmn_6A_COMB/seq_US-09-032-894-26

Seq_documentation_block:
; Sequence 26, Application US/09032894
; Patent No. 6130041
; GENERAL INFORMATION:
;   APPLICANT: Acton, Susan L.
;   TITLE OF INVENTION: SR-BI NUCLEIC ACIDS AND USES THEREFOR
;   FILE REFERENCE: MIA-005/03
;   CURRENT APPLICATION NUMBER: US/09/032,894
;   CURRENT FILING DATE: 1998-02-27
;   EARLIER APPLICATION NUMBER: 08/890,980
;   EARLIER FILING DATE: 1997-07-10
;   NUMBER OF SEQ ID NOS: 121
;   SOFTWARE: PatentIn Ver. 2.0
;   SEQ ID NO 26
;     LENGTH: 160
;     TYPE: DNA
;     ORGANISM: Human
US-09-032-894-26


alignment_scores:
             Quality:    6.00              Length:       6
               Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_46_54 x US-09-032-894-26/rev   ..

 Align Seq 1/1  to reverse of: US-09-032-894-26  from: 1  to: 160

        3 ProAlaThrProGluGln 8
          ||||||||||||||||||
       80 CCTGCCACTCCCGAGCAG 63

Seq_name: /cgn_2_ptodata/2_fmn_6A_COMB/seq_US-09-032-894-66

Seq_documentation_block:
; Sequence 66, Application US/09032894
; Patent No. 6130041
; GENERAL INFORMATION:
;   APPLICANT: Acton, Susan L.
;   TITLE OF INVENTION: SR-BI NUCLEIC ACIDS AND USES THEREFOR
;   FILE REFERENCE: MIA-005/03
;   CURRENT APPLICATION NUMBER: US/09/032,894
;   CURRENT FILING DATE: 1998-02-27
;   EARLIER APPLICATION NUMBER: 08/890,980
;   EARLIER FILING DATE: 1997-07-10
;   NUMBER OF SEQ ID NOS: 121
;   SOFTWARE: PatentIn Ver. 2.0
;   SEQ ID NO 66
;     LENGTH: 160
;     TYPE: DNA
;     ORGANISM: Human
US-09-032-894-66


alignment_scores:
             Quality:    6.00              Length:       6
               Ratio:   1.000                Gaps:       0
```

```
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US 09 126 945B 2 COPY 46_54 x US 09 031 894 66/rev   ..

 Align seq 1/1  to reverse of: US 09 031 894 66  from: 1  to: 160

        3 ProAlaThrProGluGln 8
          |||||||||||||||||||
       80 CCTGGCCACTCCCGAGCAG 63

seq name: /cgn2_6/ptodata/2/lina/6B_COMB.seq:US 09 031 626 26

seq documentation block:
; Sequence 26, Application US/09031626
; Patent No. 6228581
; GENERAL INFORMATION:
;  APPLICANT: Acton, Susan L.
;  APPLICANT: Ordovas, Jose M.
;  TITLE OF INVENTION: DIAGNOSTIC ASSAYS AND KITS FOR BODY MASS AND
;  TITLE OF INVENTION: CARDIOVASCULAR DISORDERS
;  FILE REFERENCE: MIA 005.04
;  CURRENT APPLICATION NUMBER: US/09/031,626
;  CURRENT FILING DATE: 1998-02-27
;  EARLIER APPLICATION NUMBER: 08/890,979
;  EARLIER FILING DATE: 1997-07-10
;  NUMBER OF SEQ ID NOS: 121
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
;   LENGTH: 160
;   TYPE: DNA
;   ORGANISM: Human
US 09 031 626 26


alignment scores:
              Quality:    6.00              Length:       6
                Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US 09 126 945B 2 COPY 46_54 x US 09 031 626 26/rev

 Align seq 1/1  to reverse of: US 09 031 626 26  from: 1  to: 160

        3 ProAlaThrProGluGln 8
          |||||||||||||||||||
       80 CCTGGCCACTCCCGAGCAG 63

seq name: /cgn2_6/pt_data/2/lina_68_COMB.seq:US 09 031 626 66

seq documentation block:
; Sequence 66, Application US/09031626
; Patent No. 6228581
; GENERAL INFORMATION:
;  APPLICANT: Acton, Susan L.
;  APPLICANT: Ordovas, Jose M.
;  TITLE OF INVENTION: DIAGNOSTIC ASSAYS AND KITS FOR BODY MASS AND
;  TITLE OF INVENTION: CARDIOVASCULAR DISORDERS
;  FILE REFERENCE: MIA 005.04
;  CURRENT APPLICATION NUMBER: US/09/031,626
;  CURRENT FILING DATE: 1998-02-27
;  EARLIER APPLICATION NUMBER: 08/890,979
;  EARLIER FILING DATE: 1997-07-10
;  NUMBER OF SEQ ID NOS: 121
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 66
;   LENGTH: 160
;   TYPE: DNA
;   ORGANISM: Human
US 09 031 626 66
```

```
alignment scores:
              Quality:    6.00              Length:       6
                Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US 09-126-945B 2_COPY 46_54 x US-09-031-626-66/rev   ..

 Align seq 1/1  to reverse of: US-09-031-626-66  from: 1  to: 160

        3 ProAlaThrProGluGln 8
          |||||||||||||||||||
       80 CCTGGCCACTCCCGAGCAG 63

 Sequence  [illegible] Application US/08928692 Patent No. 5958727 GENERAL INFORMATION: APPLICANT: Brody, Howard APPLICANT: Yaver, Deborah S. APPLICANT: Lamsa, Michael APPLICANT: Hansen, Kim TITLE OF INVENTION: Methods for Modifying the Production of TITLE OF INVENTION: a Polypeptide NUMBER OF SEQUENCES: 80 CORRESPONDENCE ADDRESS: ADDRESSEE: No. 5958727c No. 5958727disk of No. 5958727th America, Inc. STREET: 405 Lexington Avenue CITY: New York STATE: NY COUNTRY: USA ZIP: 10174 COMPUTER READABLE FORM: MEDIUM TYPE: Diskette COMPUTER: IBM Compatible OPERATING SYSTEM: DOS SOFTWARE: FastSEQ for Windows Version 2.0 CURRENT APPLICATION DATA: APPLICATION NUMBER: US/08/928,692 FILING DATE: 12 SEPT 1997 CLASSIFICATION: 435 ATTORNEY/AGENT INFORMATION: NAME: Lambiris, Elias J REGISTRATION NUMBER: 33,728 REFERENCE/DOCKET NUMBER: 4944.200 US TELECOMMUNICATION INFORMATION: TELEPHONE: 212-867-0123 TELEFAX: 212 808 9655 INFORMATION FOR SEQ ID NO: 84: SEQUENCE CHARACTERISTICS: LENGTH: 346 base pairs TYPE: nucleic acid STRANDEDNESS: single TOPOLOGY: linear
US 08 928 692 84


alignment scores:
              Quality:    6.00              Length:       6
                Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US 09-126-945B 2_COPY 46_54 x US 08 928 692 84/rev   ..

 Align seq 1/1  to reverse of: US 08 928 692 84  from: 1  to: 346

        1 SerProProAlaThrPro 6
          ||||||||||||||||||
      150 TCACCGCCTGCAACACCA 133
```

```
seq_name: us-09-126-945b-2_copy_46_54 seq ID: US-08-997-080-79

seq_documentation_block:
; Sequence 79, Application US/08997080
; Patent No. 5968524
;  GENERAL INFORMATION:
;     APPLICANT:  WATSON, JAMES D.
;     APPLICANT:  TAN, PAUL L.J.
;     TITLE OF INVENTION:  METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY
;     NUMBER OF SEQUENCES:  194
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Law Offices of Ann W. Speckman
;       STREET:  2601 Elliott Avenue, Suite 4185
;       CITY:  Seattle
;       STATE:  WA
;       COUNTRY:  USA
;       ZIP:  98121
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/997,080
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Sleath, Janet
;       REGISTRATION NUMBER:  37,007
;       REFERENCE/DOCKET NUMBER:  11000.1007
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  206-269-0565
;       TELEFAX:  206-269-0563
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  79:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  360 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  Genomic DNA
US-08-997-080-79


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_46_54 x US-08-997-080-79     .

 Align seq 1/1  to: US-08-997-080-79  from: 1  to: 360

        1 SerProProAlaThrPro 6
          ||||||||||||||||||
      257 TCTCCGGCCGGCGACACCC 274

seq_name: us-08-997-362-79 seq ID: US-08997362

seq_documentation_block:
; Sequence 79, Application US/08997362
; Patent No. 5985287
;  GENERAL INFORMATION:
;     APPLICANT:  Tan, Paul
;     APPLICANT:  Hiyama, Jun
;     APPLICANT:  Visser, Elizabeth
;     APPLICANT:  Skinner, Margot
;     APPLICANT:  Scott, Linda
;     APPLICANT:  Prestidge, Ross
;     TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR
;     TITLE OF INVENTION:  TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
;     NUMBER OF SEQUENCES:  194
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Law Offices of Ann W. Speckman
;       STREET:  2601 Elliott Avenue, Suite 4185
;       CITY:  Seattle
;       STATE:  WA
;       COUNTRY:  USA
;       ZIP:  98121
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/997,362
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  U.S. Patent Application No. 08/873,970
;       FILING DATE:  June 12, 1997
;       APPLICATION NUMBER:  U.S. Patent Application No. 08/287,447
;       FILING DATE:  August 29, 1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Sleath, Janet
;       REGISTRATION NUMBER:  37,007
;       REFERENCE/DOCKET NUMBER:  11000.100202
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  206-269-0565
;       TELEFAX:  206-269-0563
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  79:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  360 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  Genomic DNA
US-08-997-362-79


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_46_54 x US-08-997-362-79     .

 Align seq 1/1  to: US-08-997-362-79  from: 1  to: 360

        1 SerProProAlaThrPro 6
          ||||||||||||||||||
      257 TCTCCGGCCGGCGACACCC 274

seq_name: us-08-873-970-79 seq ID: US-08873970

seq_documentation_block:
; Sequence 79, Application US/08873970
; Patent No. 6001361
;  GENERAL INFORMATION:
;     APPLICANT:  Tan, Paul
;     APPLICANT:  Hiyama, Jun
;     APPLICANT:  Visser, Elizabeth
;     APPLICANT:  Skinner, Margot
;     APPLICANT:  Scott, Linda
;     APPLICANT:  Prestidge, Ross
;     TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR
;     TITLE OF INVENTION:  TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
;     NUMBER OF SEQUENCES:  106
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Law Offices of Ann W. Speckman
```

```
;       STREET:  2601 Elliott Avenue, Suite 4185
;       CITY:  Seattle
;       STATE:  WA
;       COUNTRY:  USA
;       ZIP:  98121
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/873,970
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/705,347
;       FILING DATE:  29-AUG-1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Sheath, Janet
;       REGISTRATION NUMBER:  37,007
;       REFERENCE/DOCKET NUMBER:  11000.100201
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  206 269 0565
;       TELEFAX:  206 269 0563
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  79:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  360 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  Genomic DNA
US 08 873 970-79


alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US 09 126 945B 2 COPY 46 54 x US-08-873-970-79   ..

 Align seq 1/1  to: US 08 873 970 79  from: 1  to: 360

        1 SerProProAlaThrPro 6
          ||||||||||||||||||
      257 TCTCCGCCCGCCACACCC 274

Seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US 09 095 855 79

seq documentation block:
; Sequence 79, Application US/09095855
; Patent No. 6160093
;  GENERAL INFORMATION:
;     APPLICANT:  Tan, Paul
;     APPLICANT:  Visser, Elizabeth
;     APPLICANT:  Skinner, Margot
;     APPLICANT:  Prestidge, Ross
;     TITLE OF INVENTION:  Compounds and Methods for
;     TITLE OF INVENTION:  Treatment and Diagnosis of Mycobacterial Infections
;     NUMBER OF SEQUENCES:  208
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Law Offices of Ann W. Speckman
;       STREET:  2601 Elliott Avenue, Suite 4185
;       CITY:  Seattle
;       STATE:  WA
;       COUNTRY:  USA
;       ZIP:  98121
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/095,855
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/705,347
;       FILING DATE:  29-AUG-1996
;       APPLICATION NUMBER:  08/873,970
;       FILING DATE:  12-JUN-1997
;       APPLICATION NUMBER:  08/997,362
;       FILING DATE:  23-DEC-1997
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Sheath, Janet
;       REGISTRATION NUMBER:  37,007
;       REFERENCE/DOCKET NUMBER:  11000.100203
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  206-269-0565
;       TELEFAX:  206 269 0563
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  79:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  360 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  Genomic DNA
US 09 095-855 79


alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US 09 126 945B 2_COPY_46_54 x US-09 095-855 79   ..

 Align seq 1/1  to: US 09 095 855 79  from: 1  to: 360

        1 SerProProAlaThrPro 6
          ||||||||||||||||||
      257 TCTCCGCCCGCCACACCC 274

Seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US 08 890 980 9

seq documentation block:
; Sequence 9, Application US/08890980
; Patent No. 5998141
;  GENERAL INFORMATION:
;     APPLICANT:  Acton, Susan L.
;     TITLE OF INVENTION:  SR-BI NUCLEIC ACIDS AND USES THEREFOR
;     NUMBER OF SEQUENCES:  86
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  FOLEY, HOAG & ELIOT LLP
;       STREET:  One Post Office Square
;       CITY:  Boston
;       STATE:  MA
;       COUNTRY:  USA
;       ZIP:  02109-2170
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  Patentin Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/890,980
;       FILING DATE:  10-JUL-1997
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Arnold, Beth E.
;       REGISTRATION NUMBER:  35,430
;       REFERENCE/DOCKET NUMBER:  MIA 005.01
```

```
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617-832-1000
;       TELEFAX:  617-832-7000
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  416 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-890-980-9


alignment_scores:
             Quality:    6.00              Length:       6
               Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_46_54 x US-08-890-980-9/rev   ..

 Align seq 1/1  to reverse of: US-08-890-980-9  from: 1  to: 416

        3 ProAlaThrProGluGln 8
          ||||||||||||||||||
      336 CCTGCCACTCCCCAGCAG 319
```

```
M of: US-09-126-945B-2_COPY_46_54 to: EST:*  Out format : pir

Date: Sep 14, 2001  7:47 PM

About: Results were produced by the GenCore software, version 4.5.
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL frame+_p2n.model -DEV xlp
 -Q /cgn2_7/tmp/US-09-126-945B-2_COPY_46_54.qry_34*/*_1.*/*11_*/*** /app_query.fasta 1 1267
 -DB=EST -QFMT fastap -SUFFIX=olip2n.rst -GAPOP=4.500
 -GAPEXT 0.050 -MINMATCH 0.100 -LOOPCL 0.000  L=*EXT 0.000
 -QGAPOP 4.500 -QGAPEXT 0.050 -XGAPOP=60.000 -XGAPEXT 60.000
 -FGAPOP 6.000 -FGAPEXT=7.000 -YGAPOP 60.000 -YGAPEXT 60.000
 -DELOP 6.000 -DELEXT=7.000 -START 1 -MATRIX olig
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE quality
 -THR_MIN=1 -ALIGN 15 -MODE LOCAL -OUTFMT pir  NORM ext  MINLEN 0
 -MAXLEN 2000000000 -USER=US09126945_2CGN1_1_4645 -NCPU=6
 -ICPU=3 -LONG[=0  -NO_XLPXY -WAIT -THREADS 1

Search Information Block:
Query: US-09-126-945B-2_COPY_46_54
Query length: 9
Database: EST:*
Database sequences: 10228115
Database length: 4314594S4
Search time (sec): 9087.630000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000


Score_list:
Sequence              Strd Orig    ZScore    EScore Len  ! Documentation ..
qb_est24:AI685592     +     9.00    167.85      2.84     386  ! AI685592 wj22e07.x1 NCI_CGAP_Pr
qb_est20:AI451780     +     9.00    165.13      4.44     544  ! AI451780 wh22f07.y1 Soares_mamm
qb_est54:AW988753     +     9.00    164.31      4.93     618  ! AW988753 qf20f07.y1 Soares_mamm
qb_est89:BF535871     +     9.00    164.12      5.05     637  ! BF535871 602051484F1 NCI_CGAP_S
qb_est76:BE563288     +     9.00    161.66      6.93     933  ! BE563288 601335410F1 NIH_MGC_9
qb_est89:BF529428     +     9.00    161.53      7.05     952  ! BF529428 602049614F1 NCI_CGAP_G
qb_est97:BG179568     +     9.00    161.27      7.28     991  ! BG179568 602382182F1 NIH_MGC_9
qb_est89:BF579116     +     9.00    156.54     13.36    2064  ! BF579116 602096058F1 NCI_CGAP_
qb_est82:BE994318     +     7.00    138.90    128.23     122  ! BE994318 QI M-030p-b6j-c-11-0-
qb_est81:BE932859     +     7.00    137.70    149.71     147  ! BE932859 CM2-HT0871-160800-427
qb_est44:AW175906     +     7.00    136.65    171.27     173  ! AW175906 qV0-BT0074-250899-004
qb_est81:BE933171     +     7.00    135.84    189.89     196  ! BE933171 MR0-HT0979-210800-022
qb_est86:BF435997     +     7.00    135.34    202.61     212  ! BF435997 CM1-CT0487-260600-573
qb_est45:AW446128     -     7.00    134.43    227.52     244  ! AW446128 27304 MARC 4B-V Bos ta
qb_est68:BB598238     -     7.00    134.17    235.26     254  ! BB598238 BB598238 RIKEN full le
qb_est33:AV159382     -     7.00    133.43    258.25     285  ! AV159382 AV159382 Mus musculus
qb_est52:AW853811     -     7.00    133.34    261.74     289  ! AW853811 RC3-CT0254-120200-025
qb_est54:BB002961     -     7.00    133.17    267.72     297  ! BB002961 BB002961 RIKEN full le
qb_est97:BG152666     -     7.00    132.73    284.27     318  ! BG152666 ta58q05.x1 NCI_CGAP_B
qb_est86:BF457144     -     7.00    132.00    310.97     356  ! BF457144 MR0-HT0923-250800-301
qb_est41:AV749644     -     7.00    131.86    316.73     364  ! AV749644 AV749644 NPC Homo sapi
qb_est86:BF457141     -     7.00    131.66    324.62     375  ! BF457141 MR0-HT0923-250800-301
qb_est86:BF457136     -     7.00    131.66    324.62     375  ! BF457136 MR0-HT0923-250800-301
qb_est51:AW763829     -     7.00    131.44    333.09     388  ! AW763829 qr66c02.x1 NCI_CGAP_Ma
qb_est72:BE401399     -     7.00    131.23    341.12     401  ! BE401399 bb72d01.x1 NIH_MGC_12
qb_est27:AI969060     -     7.00    130.96    355.09     418  ! AI969060 wq70d12.x1 NCI_CGAP_G
qb_est94:BF931752     -     7.00    130.92    357.20     421  ! BF931752 CM2-NT0210-141200-624
qb_est13:AW399421     -     7.00    130.86    360.68     425  ! AW399421 QI H-B11-adt-a-12-0-B1
qb_est22:AI632183     -     7.00    130.56    373.95     445  ! AI632183 ts85h07.x1 NCI_CGAP_G
qb_est42:AW081365     -     7.00    130.46    378.88     452  ! AW081365 xc41f06.x1 NCI_CGAP_G
qb_est78:BE724045     -     7.00    130.45    379.49     453  ! BE724045 198717 MARC 4B0V Bos t
qb_est22:AI573182     -     7.00    130.42    380.88     455  ! AI573182 tt70f01.x1 NCI_CGAP_P
qb_est24:AI744752     -     7.00    130.42    380.88     455  ! AI744752 tt14e07.x1 NCI_CGAP_P
qb_est25:AI806249     -     7.00    130.40    386.40     463  ! AI806249 wt26b09.x1 Soares_NFL
qb_est92:BF763009     -     7.00    130.29    387.09     464  ! BF763009 IL2-CN0048-230300-207
qb_est107:R28490      -     7.00    130.18    392.60     472  ! R28490 yh54e12.s1 Soares placen
qb_est39:AV615792     -     7.00    129.94    404.94     490  ! AV615792 AV615792 Bos taurus cD
qb_est77:BE683391     -     7.00    129.94    404.94     490  ! BE683391 182838 MARC 4B0V Bos t
qb_ssa81:AZ715454     -     7.00    129.16    447.51     553  ! AZ715454 RPCI-24_71N4.TV RPCI-2
qb_est84:BF073843     -     7.00    129.08    452.18     560  ! BF073843 IL2-CT0242-080200-G2
qb_est31:AA159866     -     7.00    129.03    454.85     564  ! AA159866 AA159866 yz64a11.r1 Hom
qb_est96:BG087394     -     7.00    128.96    462.84     576  ! BG087394 HS1594A09-S NIA Mouse
qb_est48:AW555629     -     7.00    128.28    501.03     634  ! AW555629 602577B0F1 NIH_MGC_8 Mouse Ne
qb_est3:AA197632      -     7.00    128.21    505.60     641  ! AA197632 mn22a05.r1 Soares_t
qb_est17:AI225303     -     7.00    128.19    506.90     643  ! AI225303 ne06b09.y1 Soares_N

seq_name: qb_est24:AI685592

seq_documentation_block:
LOCUS       AI685592              386 bp    mRNA            EST       15-DEC-1999
DEFINITION  wj22e07.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2251596 3'
            similar to contains PTR5.t3 MSR1 repetitive element ;, mRNA
            sequence.
ACCESSION   AI685592
VERSION     AI685592.1  GI:4896886
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 386)
  AUTHORS   NCI CGAP Pr28.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael K.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www.bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1143  Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 480.
FEATURES             Location/Qualifiers
     source          1..386
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2251596"
                     /clone_lib="NCI_CGAP_Pr28"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT7T3-Pac (Pharmacia)
                     with a modified polylinker; Plasmid DNA from the
                     normalized library NCI_CGAP_Pr22 was prepared, and ss
                     circles were made in vitro. Following BAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR amplified cDNAs from a pool
                     of 5,000 clones made from the same library (clones
                     986608-986759, 1101192-1101959, and 1217928-1220915).
                     Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT        72 a    136 c    110 g     64 t
ORIGIN


alignment_scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_46_54 x AI685592  ..

 Align Seq 1/1  to: AI685592  from: 1  to: 386

       1 SerProProAlaThrProGluGlnGly 9
         ||||||||||||||||||||||||||
```

```
     296 AGTCCACCCGCCACGCCCGAGCAGGGC 322

seq_name: qb_est20:AI451780

seq_documentation_block:
LOCUS       AI451780       544 bp    mRNA            EST       15-MAR-2000
DEFINITION  mb22l07.y1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
            IMAGE:430181 5', mRNA sequence.
ACCESSION   AI451780
VERSION     AI451780.1  GI:4306493
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 544)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,L.,
            Underwood,K., Steptoe,M., Pepin,K., Allen,M., Bowers,Y., Berry
            ,R., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Marra M/WashU NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, Mo 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            This read is a FEATURE-PREDICTED sequence [ ]
            This read has been verified (found to hit its original self in the
            correct orientation)
            MGI:211581
            Seq primer:  40RP from Gibco
            High quality sequence stop: 462
            POLYA No.
FEATURES             Location/Qualifiers
     source          1..544
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:430181"
                     /clone_lib="Soares mouse p3NMF19.5"
                     /dev_stage="19.5 dpc total fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M Fatima Bonaldo. RNA was kindly provided by
                     Dr. Minoru Ko (Wayne State University)."
BASE COUNT       104 a    190 c    153 g     97 t
ORIGIN


alignment_scores:
            Quality:    9.00            Length:       9
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 DS 09-126-945B-2_COPY 46-54 x AI451780  ..

 Align seq 1/1  to: AI451780  from: 1  to: 544

       1 SerProProAlaThrProGluGlnGly 9
         |||||||||||||||||||||||||||
     538 AGTCCACCCGCCACGCCCGAGCAGGGC 564

seq_name: qb_est89:BF535871

seq_documentation_block:
LOCUS       BF535871       637 bp    mRNA            EST       11-DEC-2000
DEFINITION  602051384F1 NCI_CGAP_Sk2 Mus musculus cDNA clone IMAGE:4190637 5',
            mRNA sequence.
ACCESSION   BF535871
VERSION     BF535871.1  GI:11623239
KEYWORDS    EST.
SOURCE      house mouse.
```

```
     466 AGTCCACCCGCCACGCCCGAGCAGGGC 492

seq_name: qb_est54:AW988753

seq_documentation_block:
LOCUS       AW988753       618 bp    mRNA            EST       02-JUN-2000
DEFINITION  wi28h07.y1 Soares mammary gland NMLMG Mus musculus cDNA clone
            IMAGE:1511941 5' similar to TR:O95238 O95238 EIS TRANSCRIPTION
            FACTOR PDEF. ;, mRNA sequence.
ACCESSION   AW988753
VERSION     AW988753.1  GI:8183979
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 618)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbcr@mail.nih.gov
            This clone is available royalty free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:938793
            Seq primer:  40RP from Gibco
            High quality sequence stop: 407.
FEATURES             Location/Qualifiers
     source          1..618
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1511941"
                     /clone_lib="Soares mammary gland NMLMG"
                     /sex="female (lactating)"
                     /tissue_type="mammary gland"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from mammary
                     gland tissue from a lactating female, and was then primed
                     with a Not I - oligo(dT) primer. Double stranded  cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I  and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector.  Library is normalized.  Library
                     was constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       114 a    225 c    164 g    115 t
ORIGIN


alignment_scores:
            Quality:    9.00            Length:       9
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 DS 09-126-945B-2_COPY 46-54 x AW988753  ..

 Align seq 1/1  to: AW988753  from: 1  to: 618

       1 SerProProAlaThrProGluGlnGly 9
         |||||||||||||||||||||||||||
     538 AGTCCACCCGCCACGCCCGAGCAGGGC 564

seq_name: qb_est89:BF535871

seq_documentation_block:
LOCUS       BF535871       637 bp    mRNA            EST       11-DEC-2000
DEFINITION  602051384F1 NCI_CGAP_Sk2 Mus musculus cDNA clone IMAGE:4190637 5',
            mRNA sequence.
ACCESSION   BF535871
VERSION     BF535871.1  GI:11623239
KEYWORDS    EST.
SOURCE      house mouse.
```

```
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorphi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 647)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9518  row: q  column: 22
            High quality sequence stop: 647.
FEATURES             Location/Qualifiers
     source          1..647
                     /organism "Mus musculus"
                     /strain "FVB/N"
                     /db_xref="taxon:10090"
                     /clone "IMAGE:4190647"
                     /clone_lib "NCI_CGAP_SG2"
                     /lab_host "DH10B (ff phage-resistant)"
                     /note "Organ: salivary gland; Vector: pSPORT1; Site 1:
                     NotI; Site_2: SalI; Cloned unidirectionally, Primer: Oligo
                     dT  Average insert size 1.8 kb, Constructed by Life
                     Technologies, Note: this is a NCI_CGAP Library."
BASE COUNT       120 a    219 c    179 g    114 t      5 others
ORIGIN


alignment_scores:
            Quality:    9.00             Length:      9
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_46_54 x BE535871   ..

 Align seq 1/1  to: BE535871  from: 1  to: 647

       1 SerProProAlaThrProGluGlnGly 9
         |||||||||||||||||||||||||||
     384 AGTCCACCCGCCACCCCTGAGCAGGGC 409

seq_name: qb_est76:BE563288

seq_documentation_block:
LOCUS       BE563288       933 bp    mRNA            EST       15-AUG-2000
DEFINITION  601335410F1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:3689402 5',
            mRNA sequence.
ACCESSION   BE563288
VERSION     BE563288.1  GI:9807100
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 933)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9514  row: k  column: 03
            High quality sequence stop: 678.
FEATURES             Location/Qualifiers
     source          1..933
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone "IMAGE:3689402"
                     /clone_lib "NIH_MGC_39"
                     /tissue_type "adenocarcinoma"
                     /lab_host "DH10B (phage-resistant)"
                     /note "Organ: pancreas; Vector: pSPORT; Site 1: AbaI;
                     Site_2: EcoRI; cDNA made by oligodT primed,
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed
                     by Lisa Heda in the laboratory of Gerald M. Rubin
                     (University of California, Berkeley) using ZAP cDNA
                     synthesis kit (Stratagene) and Superscript II RT (Life
                     Technologies)."
BASE COUNT       148 a    316 c    319 g    150 t
ORIGIN


alignment_scores:
            Quality:    9.00             Length:      9
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_46_54 x BE563288   ..

 Align seq 1/1  to: BE563288  from: 1  to: 933

       1 SerProProAlaThrProGluGlnGly 9
         |||||||||||||||||||||||||||
     531 AGTCCACCCGCCACCCCTGAGCAGGGC 557

seq_name: qb_est89:BE539428

seq_documentation_block:
LOCUS       BE539428       952 bp    mRNA            EST       11-DEC-2000
DEFINITION  602049614F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4189178 5',
            mRNA sequence.
ACCESSION   BE539428
VERSION     BE539428.1  GI:11626809
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorphi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 952)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9514  row: k  column: 03
            High quality sequence stop: 646.
FEATURES             Location/Qualifiers
     source          1..952
                     /organism "Mus musculus"
                     /strain "FVB/N"
                     /db_xref "taxon:10090"
                     /clone "IMAGE:4189178"
                     /clone_lib "NCI_CGAP_SG2"
                     /lab_host "DH10B (ff phage-resistant)"
```

```
                     /note "Organ: salivary gland; Vector: pCMV SPORT6; Site 1:
                     Not1; Site 2: Sal1; Cloned unidirectionally. Primer: oligo
                     dT. Average insert size 1.8 kb. Constructed by Life
                     Technologies. Note: this is a NCI CGAP Library."
BASE COUNT       183 a    312 c    298 g    198 t      1 others
ORIGIN


alignment scores:
            Quality:     9.00            Length:       9
              Ratio:    1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_46_54 x BE539428    ..

 Align seq 1/1  to: BE539428  from: 1  to: 952

       1 SerProProAlaThrProGluGlnGly 9
         |||||||||||||||||||||||||||
     925 AGTCCACCCGCCACCCCTGAGCAGGGC 954

Seq name: qb_est37:BG179568

seq documentation block:
LOCUS        BG179568       991 bp    mRNA            EST       06-FEB-2001
DEFINITION  602328182F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4429413 5',
            mRNA sequence.
ACCESSION   BG179568
VERSION     BG179568.1  GI:12586221
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 991)
  AUTHORS   NIH_MGC http://mgc.nci.nih.gov/
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10181  row: d  column: 22
            High quality sequence stop: 655.
FEATURES             Location/Qualifiers
     source          1..991
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4429413"
                     /clone_lib="NIH_MGC_91"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage resistant)"
                     /note="Organ: prostate; Vector: pCMV SPORT6; Site_1: Not1;
                     Site_2: Sal1; Cloned unidirectionally; oligo dT primed.
                     Average insert size 1.4 kb. Library enriched for
                     full length clones and constructed by Life Technologies.
                     Note: this is a NIH MGC Library."
BASE COUNT       190 a    351 c    305 g    145 t
ORIGIN


alignment scores:
            Quality:     9.00            Length:       9
              Ratio:    1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
```

```
 US-09-126-945B-2_COPY_46_54 x BG179568    ..

 Align seq 1/1  to: BG179568  from: 1  to: 991

       1 SerProProAlaThrProGluGlnGly 9
         |||||||||||||||||||||||||||
     39 AGTCCACCCGCCACCCCTGAGCAGGGC 65

Seq name: qb_est39:BE539116

seq documentation block:
LOCUS        BE539116      2064 bp    mRNA            EST       12-DEC-2000
DEFINITION  602096058F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4215901 5',
            mRNA sequence.
ACCESSION   BE539116
VERSION     BE539116.1  GI:11652828
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2064)
  AUTHORS   NIH_MGC http://mgc.nci.nih.gov/
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9792  row: d  column: 14
            High quality sequence stop: 645.
FEATURES             Location/Qualifiers
     source          1..2064
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4215901"
                     /clone_lib="NCI_CGAP_Co24"
                     /lab_host="DH10B (T1 phage resistant)"
                     /note="Organ: colon; Vector: pCMV SPORT6; Site 1: Not1;
                     Site 2: Sal1; Cloned unidirectionally.  Primer: oligo dT.
                     Average insert size 1.6 kb. Constructed by Life
                     Technologies. Note: this is a NCI CGAP Library."
BASE COUNT       624 a    621 c    620 g    199 t
ORIGIN


alignment scores:
            Quality:     9.00            Length:       9
              Ratio:    1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_46_54 x BE539116    ..

 Align seq 1/2  to: BE539116  from: 1  to: 2064

       1 SerProProAlaThrProGluGlnGly 9
         |||||||||||||||||||||||||||
     121 AGTCCACCCGCCACCCCTGAGCAGGGC 147

Seq name: qb_est182:BE994418

seq documentation block:
LOCUS        BE994418       122 bp    mRNA            EST       05-OCT-2000
DEFINITION  UI_M_CG0p_bi1_e_11_g_BI.s1 NIH_BMAP_Rel4_32 Mus musculus cDNA clone
            UI_M_CG0p_bi1_e_11_0_UI 3', mRNA sequence.
ACCESSION   BE994418
```

```
VERSION     BE994318.1  GI:10677467
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 122)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            5001 Executive Blvd. Room 7N-719, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            EMail: mESTmail.nih.gov
            Oligo-dT track not found, Not 1 site shown in beginning of sequence
            is likely internal to the message. DNA Library is part of M.B.
            Soares Lab Clone distribution: Researchers may obtain BMAP cDNA
            clones from RESEARCH GENETICS. It should be noted that Bonaldo Soares
            is generating a small number of additional specialized
            non-redundant arrays of BMAP cDNAs whose availability will be
            considered under appropriate and limited collaborative arrangements
            Seq primer: M13 Forward
            POLYA No.
FEATURES             Location/Qualifiers
     source          1..122
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:[illegible]"
                     /clone_lib="NIH_BMAP_Rot4_S2"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: p2714-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_Rot4_S2 library is a subtracted library,
                     ultimately derived from mouse retina tissue libraries at
                     various stages of development. For a detailed description
                     of the library from which this clone was derived, please
                     visit our web site at brainest.enq.uiowa.edu.
                     IAG_SEQ: None found"
BASE COUNT       21 a      46 c      40 g      25 t
ORIGIN


alignment scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_46_54 x BE994318    ..

 Align seq 1/1  to: BE994318  from: 1  to: 122

        2 ProProAlaThrProGluGln 8
          ||||||||||||||||||||||
       81 CCTCCGGCGACTCCAGAGCAG 101

seq_name: qb_est81:BE932850

seq_documentation_block:
LOCUS       BE932850       147 bp    mRNA             EST       02-OCT-2000
DEFINITION  CM2-HT0871-160800-327-d11 HT0871 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE932850
VERSION     BE932850.1  GI:10458926
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 147)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,L.H.,
            Brumstein,A., de Oliveira,P.S., Bucher,P., Jonquiel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20204863
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the EMBL/GSB Human Cancer Genome
            Project. This entry can be seen in the following URL
            [http://www.ludwig.org.br/scripts/getitml2.pl?id_seq=HT0871 160
            800 327 d11&t4 2000-08-16&t4 1)
            Seq primer: puc18 forward
            High quality sequence start: 25
            High quality sequence stop: 147.
FEATURES             Location/Qualifiers
     source          1..147
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="HT0871"
                     /dev_stage="Adult"
                     /note="organ: head_neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 136,716    Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       26 a      52 c      31 g      38 t
ORIGIN


alignment scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_46_54 x BE932850    ..

 Align seq 1/1  to: BE932850  from: 1  to: 147

       1 SerProProAlaThrProGlu 7
         |||||||||||||||||||
      68 TCTCCTCCAGCCACTCCTGAG 88

seq_name: qb_est43:AW175906

seq_documentation_block:
LOCUS       AW175906       173 bp    mRNA             EST       16-NOV-1999
DEFINITION  oy0 BT0874 253899 894 d04 BT0874 Homo sapiens cDNA, mRNA sequence
ACCESSION   AW175906
VERSION     AW175906.1  GI:6441944
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 173)
  AUTHORS   HCGP http://www.ludwig.org.br/ORESTES.
  TITLE     The EMBL/GSB Human Cancer Genome Project
```

```
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo SP,
            Brazil
            Tel: +55 11 2704922
            Fax: +55 11 2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/getbamI2.pl?I1 gV0&I2 gV0 BT0074
            250899 004 d04&I3 1999 08:25&I4 1)
            Seq primer: puc 18 forward
            High quality sequence start: 73
            High quality sequence stop: 173.
FEATURES             Location/Qualifiers
     source          1..173
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="BT0074"
                     /dev_stage="Adult"
                     /note="organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT       29 a      59 c      57 g      28 t
ORIGIN


Alignment scores:
            Quality:     7.00              Length:       7
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-126-945B-2_COPY_46_54 x AW175906    ..

 Align seq 1/1   to: AW175906   from: 1   to: 173

        1 SerProProAlaThrProGlu 7
          |||||||||||||||||||||
      90 TCTCCACCTGCCACCCCGGAA 110

Seq name: qb_est81:BE933171

Seq documentation block:
LOCUS       BE933171       196 bp    mRNA            EST       02-OCT-2000
DEFINITION  RC1 BT0879-210800 021-c12 BT0879 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE933171
VERSION     BE933171.1  GI:10459247
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 196)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Buchet,P., Jonquneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo SP,
            Brazil
            Tel: +55 11 2704922
            Fax: +55 11 2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/getbamI2.pl?I1 gV0&I2 gV0 BT0879
            800-021-c12&I3 2000 08:21&I4 1)
            Seq primer: puc 18 forward
            High quality sequence start: 22
            High quality sequence stop: 196.
FEATURES             Location/Qualifiers
     source          1..196
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="BT0879"
                     /dev_stage="Adult"
                     /note="organ: head_neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       57 a      70 c      57 g      32 t
ORIGIN


Alignment scores:
            Quality:     7.00              Length:       7
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-126-945B-2_COPY_46_54 x BE933171    ..

 Align seq 1/1   to: BE933171   from: 1   to: 196

        1 SerProProAlaThrProGlu 7
          |||||||||||||||||||||
       18 TCTCCACCTGCCACCCCAGAG 38

Seq name: qb_est86:BE935997

Seq documentation block:
LOCUS       BE935997       212 bp    mRNA            EST       22-NOV-2000
DEFINITION  CM4 C10487 260900-574-D05 CT0487 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE935997
VERSION     BE935997.1  GI:11306745
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 212)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Buchet,P., Jonquneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo SP,
            Brazil
            Tel: +55 11 2704922
            Fax: +55 11 2707001
```

```
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/cgi-bin/ESTlib.pl?lib=CM1&id=CM1
            260900-574-D05&t3-2000-09-26&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 13
            High quality sequence stop: 211.
FEATURES             Location/Qualifiers
     source          1. .212
                     /organism="Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone_lib "CT0487"
                     /dev_stage "Adult"
                     /note "Organ: colon; Vector: pucl8; Site 1: SmaI; Site 2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT        49 a      52 c      75 g      36 t
ORIGIN


alignment_scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_46_54 x BF335997/rev    ..

 Align seq 1/1  to reverse of: BF335997  from: 1  to: 212

       1 SerProProAlaThrProGlu 7
         |||||||||||||||||||||
      81 TCCCCGCCTGCAACCCCTGAG 61

seq_name: gb_est45:AW346128

seq_documentation_block:
LOCUS        AW346128        244 bp    mRNA            EST       14-JUL-2000
DEFINITION   27304 MARC 4B0V Bos taurus cDNA 5', mRNA sequence.
ACCESSION    AW346128
VERSION      AW346128.1  GI:6843838
KEYWORDS     EST.
SOURCE       cow.
  ORGANISM   Bos taurus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
             Bovidae; Bovinae; Bos.
REFERENCE    1  (bases 1 to 244)
  AUTHORS    Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
             Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
             ,W.W. and Keele,J.W.
  TITLE      Design and use of four pooled tissue normalized cDNA libraries for
             EST discovery in cattle
  JOURNAL    Unpublished (2000)
COMMENT      Contact: Smith TPL
             USDA, ARS, US Meat Animal Research Center
             PO Box 166, Clay Center, NE 68933-0166, USA
             Tel: 402 762 4466
             Fax: 402 762 4390
             Email: smith@email.marc.usda.gov
             Single pass sequencing. Bases called and trimmed with phred
             v0.980904.e. Vector identified by cross_match with the -minscore 20
             and -minmatch 12 options.
             PCR PRimers
             FORWARD: AGGAAACAGCTATGACCAT
             BACKWARD: GTTTTCCCAGTCACGACG
             Plate: 18  row: J  column: 21
             Seq primer: ATTTAGGTGACACTATAG.

FEATURES             Location/Qualifiers
     source          1. .244
                     /organism "Bos taurus"
                     /db_xref "taxon:9913"
                     /clone_lib "MARC 4B0V"
                     /tissue_type "pooled"
                     /lab_host="DH10B"
                     /note "Vector: pCMV SPORT6; Site_1: XbaI; Site_2  XhoI;
                     Library made from pooled tissue from day 20 and day 40
                     embryos."
BASE COUNT        55 a      52 c     104 g      33 t
ORIGIN


alignment_scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_46_54 x AW346128/rev    ..

 Align seq 1/1  to reverse of: AW346128  from: 1  to: 244

       1 SerProProAlaThrProGlu 7
         |||||||||||||||||||||
     148 TCCCCGCCTGCAACCCCTGAG 128

seq_name: gb_est68:BB598238

seq_documentation_block:
LOCUS        BB598238        254 bp    mRNA            EST       07-OCT-2000
DEFINITION   BB598238 RIKEN full length enriched, adult male liver tumor Mus
             musculus cDNA clone 2730002B23 5', mRNA sequence.
ACCESSION    BB598238
VERSION      BB598238.1  GI:11506849
KEYWORDS     EST.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 254)
  AUTHORS    Aizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T., Carninci,P.
             , Hamanaki,F., Hayatsu,N., Hiraoka,T., Hirozane,T., Hodcyama,Y.,
             Imotani,F., Ishii,Y., Itoh,M., Izawa,M., Kawai,J., Kojima,Y., Kondo
             ,H., Kuenkabe,M., Matsuyama,T., Miyazaki,A., Nakamura,M., Nishi,K.,
             Nomura,K., Numazaki,F., Okazaki,Y., Okido,T., Owa,C., Sakai,C.,
             Sakai,K., Sasaki,D., Sato,K., Shibata,K., Shibata,Y., Shinagawa,A.,
             Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Tanaka
             ,T., Toya,T., Watahiki,A., Yamamura,T., Yasunishi,A., Yoshida,K.,
             Yoshiki,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE      RIKEN Mouse ESTs (Aizawa,K. et al. 2000)
  JOURNAL    Unpublished (2000)
COMMENT      Contact: Yoshihide Hayashizaki
             Genome Exploration Research Group, Life Science Tsukuba Center,
             Genome Science Laboratory
             The Institute of Physical and Chemical Research (RIKEN), Genomic
             Sciences Center
             3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
             Tel: +81-298-36-9013
             Fax: +81 298 36 9098
             Email: genome-res@rtc.riken.go.jp,
             URL: http://genome.rtc.riken.go.jp/
             Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
             ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Thermostabilization and thermoactivation of thermolabile enzymes by
             trehalose and its application for the synthesis of full length
             cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
             Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
             ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
             system. Genome Res. 9 (5), 463-470 (1999)
```

```
              Carninci,P. and Hayashizaki,Y.
              High efficiency full length cDNA cloning, Methods Enzymol. 303,
             19-44 (1999)
              Please visit our web site (http://genome.gsc.riken.go.jp) for
             further details.
FEATURES               Location/Qualifiers
     source            1..254
                       /organism "Mus musculus"
                       /db_xref "taxon:10090"
                       /clone "C730002H24"
                       /clone_lib "RIKEN full length enriched, adult male liver
                       tumor"
                       /sex "male"
                       /tissue_type "liver tumor"
                       /dev_stage "adult"
                       /lab_host "DH10B"
                       /note "Site 1: SalI; Site 2: BamHI; cDNA library was
                       prepared and sequenced in Mouse Genome Encyclopedia
                       Project of Genome Exploration Research Group in Riken
                       Genomic Sciences Center and Genome Science Laboratory in
                       RIKEN, Division of Experimental Animal Research in Riken
                       contributed to prepare mouse tissues. 1st strand cDNA was
                       primed with a primer [5'
                       GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was
                       prepared by using trehalose thermo activated reverse
                       transcriptase and subsequently enriched for full-length by
                       cap-trapper. Second strand cDNA was prepared with the
                       primer adapter of sequence [5'
                       GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3'], cDNA
                       was cleaved with BamHI and XhoI. Vector: a modified
                       pBluescript KS(+) after bulk excision from Lambda ELC I.
                       Tissue was provided by William A. Held, Roswell Park
                       Cancer Institute, Department of Molecular and Cellular
                       Biology, Elm and Carlton Streets, Buffalo, NY 14263, whose
                       assistance we gratefully acknowledge."
BASE COUNT        57 a      56 c      93 g      48 t
ORIGIN


Alignment scores:
            Quality:     7.00              Length:       7
              Ratio:   1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-126-945B-2 (COPY 46_54 x BB598248/rev  ..

 Align seq 1/1  to reverse of: BB598248  from: 1  to: 254

        2 ProProAlaThrProGluGln 8
          ||||||||||||||||||||||
       63 CCCCCGGCGACTCCGAACAA 43
```

```
QM of: US-09-126-945B-2_COPY_66_74 to: GenEmbl:*   out_format : pts

Date: Sep 14, 2001  9:36 PM

About: Results were produced by the GenCore software, version 4.5.
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL frame*_p2n.model -DEV xlp
 -Q "/tmp2_1/PSPT_sp_0_29360/query_1[1-%d].fasta" -DB GenEmbl -QFMT fastap -SUFFIX olip2n.rge -MAPOP 4.5e6
 -GAPEXT 0.050 -MINMATCH 0.100 -LOOPCL 0.000 -LOOPEXT 0.000
 -QGAPOP 4.500 -QGAPEXT 0.050 -XGAPOP 60.000 -XGAPEXT 60.000
 -FGAPOP 6.000 -FGAPEXT 7.000 -YGAPOP 60.000 -YGAPEXT 60.000
 -DELOP 6.000 -DELEXT 7.000 -START 1 -MATRIX oligo
 -FRAME human40.cdi -LIST 45 -DOCALIGN 200 -THR_SCORE quality
 -THR_MIN 1 -ALIGN 15 -MODE LOCAL -OUTFMT pts  NORM ext  MINLEN 0
 -MAXLEN 2000000000 -USER US09126945B_COPY_1_7591  NCPU 6
 -LOG 4 -LONGLOG -NO_XLPXY -WAIT  THREADS 1

Search information block:
Query: US-09-126-945B-2_COPY_66_74
Query length: 9
Database: GenEmbl:*
Database sequences: 1444157
Database length: -856060004
Search time (sec): 6503.700000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000


score_list:
Sequence                 Strd Orfa    %Score   %Score Len  ! Documentation
qb_pat1:AX048148          +     9.00  174.99   0.1645    1894  ! AX048148 Sequence 1 from Paten
qb_pr4:AF071538           +     9.00  174.99   0.1645    1894  ! AF071538 Homo sapiens Ets tran
qb_pr1:AB031549           +     9.00  174.98   0.1648    1898  ! AB031549 Homo sapiens PSE mRNA
qb_pat1:AX048159          +     9.00  171.64   0.2530    3417  ! AX048159 Sequence 12 from Pate
qb_pr6:AL157472           -     9.00  148.78   4.74      150660  ! AL157472 Human DNA sequence
qb_pl3:GMCHLH             -     7.00  129.67   55.03     4449  ! AJ891091 Glycine max mRNA for
qb_in3:DDU20132           -     7.00  127.77   70.26     5115  ! U20132 Dictyostelium discoidea
qb_ba3:SC7H2              -     7.00  116.13   312.41    42655  ! AL109732 Streptomyces coelico
qb_htq21:AL357145         +     7.00  111.52   564.23    92080  ! AL357145 Homo sapiens chromos
qb_pl3:NC17E5             +     7.00  109.45   745.60    130049  ! AL513467 Neurospora crassa D
qb_pl1:AC006423           +     7.00  109.38   742.74    131692  ! AC006423 Arabidopsis thaliar
qb_htq25:CNS05TEV         +     7.00  109.27   753.63    134214  ! AL359499 Homo sapiens chrome
qb_htq14:AC068677         +     7.00  108.01   885.71    165613  ! AC068677 Homo sapiens chrome
qb_htq22:AL390787         +     7.00  107.87   900.94    169329  ! AL390787 Homo sapiens chrom
qb_htq1:AC007608          +     7.00  107.82   906.77    170757  ! AC007608 Homo sapiens chrome
qb_htq21:AL359203         +     7.00  107.81   907.83    171015  ! AL359203 Homo sapiens chrom
qb_htq3:AC012340          -     7.00  107.71   919.96    173997  ! AC012340 Homo sapiens chrom
qb_htq11:AC027296         -     7.00  107.69   922.33    174581  ! AC027296 Homo sapiens chrome
qb_htq21:AL357501         -     7.00  107.62   930.23    176530  ! AL357501 Homo sapiens chrom
qb_htq10:AC025691         +     7.00  107.33   966.64    185576  ! AC025691 Homo sapiens chrome
qb_htq15:AC073600         +     7.00  106.95   1.0e-03   197667  ! AC073600 Mus musculus clone
qb_htq13:AC061963         +     7.00  106.82   1.0e-03   201779  ! AC061963 Mus musculus chrom
qb_htq3:AC011318          +     7.00  106.53   1.1e-03   212041  ! AC011318 Homo sapiens chrome
qb_htq1:AC007334          +     7.00  106.20   1.1e-03   223887  ! AC007334 Homo sapiens chrome
qb_pat1:AR050998          +     6.00  141.65   11.84     21  ! AR050998 Sequence 10 from patent
qb_pat1:A41885            +     6.00  130.64   48.61     132  ! A41885 Sequence 5 from Patent W
qb_pr10:HUMSCG11          +     6.00  129.05   59.57     172  ! 186416 Human cell surface glyc
qb_sts1:AU024855          -     6.00  124.20   111.87    387  ! AU024855 Rattus norvegicus, cl
qb_sts2:G16488            -     6.00  124.00   113.92    400  ! G16488 Human STS SHGC-14465 cl
qb_pr10:HSU8CHL1          -     6.00  123.58   120.22    429  ! AF076269 Homo sapiens ubiquitin
qb_pr10:HUMCD44J          +     6.00  121.97   147.73    561  ! M83328 Human cell surface glyc
qb_htq14:AC056349         -     6.00  120.60   176.07    705  ! AC056349 Giardia intestinalis c
qb_htq16:AC078765         +     6.00  120.02   189.74    777  ! AC078765 Giardia intestinalis c
qb_htq12:AC037758         +     6.00  119.57   201.07    838  ! AC037758 Giardia intestinalis c
qb_pat2:146764            +     6.00  119.14   212.40    900  ! 146764 Sequence 6 from patent U
qb_htq17:AC081168         -     6.00  119.04   215.12    915  ! AC081168 Giardia intestinalis c
qb_htq15:AC071846         +     6.00  118.91   218.90    936  ! AC071846 Giardia intestinalis c
qb_htq12:AC040437         -     6.00  118.74   223.74    963  ! AC040437 Giardia intestinalis c
qb_sts1:CNS0761A          +     6.00  118.69   224.98    970  ! AL450820 clone XBA0AB001B11 of
qb_htq13:AC050795         +     6.00  118.62   227.12    982  ! AC050795 Giardia intestinali
qb_htq12:AC030438         +     6.00  118.49   230.84    1003  ! AC030438 Giardia intestinal
qb_htq12:AC039065         +     6.00  118.49   230.84    1003  ! AC039065 Giardia intestinal
qb_htq13:AC049612         +     6.00  118.46   231.90    1009  ! AC049612 Giardia intestinal
qb_htq13:AC056348         +     6.00  118.44   232.25    1011  ! AC056348 Giardia intestinal
qb_pat1:A26465            +     6.00  118.43   232.78    1014  ! A26465 H.sapiens vCD44 resi

seq_name: qb_pat1:AX048148

seq_documentation_block:
LOCUS       AX048148      1894 bp    DNA             PAT       15-DEC-2000
DEFINITION  Sequence 1 from Patent WO0070092.
ACCESSION   AX048148
VERSION     AX048148.1  GI:11876967
KEYWORDS
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1894)
  AUTHORS   Kaufmann,J. and Xin,H.
  TITLE     Expression of ets domain proteins in cancer
  JOURNAL   Patent: WO 0070092 A 1 23-NOV-2000;
            Chiron Corporation (US)
FEATURES             Location/Qualifiers
     source          1..1894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             416..1423
                     /note="Human JKE1S."
                     /codon_start=1
                     /protein_id="CAC19002.1"
                     /db_xref="GI:11876968"
                     /translation="MQSASPGLSSVSPSHLETPETDVSPTGLEPAAAQAVQLERRDWS
                     PSPRAIPEQELSAEYLSYPDMLYPEDSSWAAKAPQASSEEPPEEEEQCEVIDSQAPA
                     GSLDLVPGGLITEEHSLEQVQSMVVGEVLKDIETACKLLNITADPMDWSPSNVQKWLL
                     WIEHQYRLPPMGKAFQELAGKELCAMSEEQFPQRSPLGGIVLHAHLDYIWKSAAWMKER
                     ISPSATRYCASTSPFSWDPSFVDSSCSQQETHIWQFTKEFLTKPDSYGREIPWELNELK
                     GIFEKIFDSAQVAFIWSIPENNPTAMNYDELSPSTPQYYEKCILRKFIDSQKLVYQEVHP
                     I"
BASE COUNT       368 a    653 c    571 g    302 t
ORIGIN


alignment_scores:
            Quality:    9.00              Length:      9
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_66_74 x AX048148  ..

 Align seq 1/1  to: AX048148  from: 1  to: 1894

        1 LeuTyrProGluAspSerSerIrpAla 9
          ||||||||||||||||||||||||||
      611 CTGTACCCGGAGGACAGCAGCTGGGCA 637

seq_name: qb_pr4:AF071538

seq_documentation_block:
LOCUS       AF071538      1894 bp    mRNA            PRI       10-JAN-2000
DEFINITION  Homo sapiens EUS transcription factor EEF (EEF) mRNA, complete
            cds.
ACCESSION   AF071538
VERSION     AF071538.1  GI:4007417
KEYWORDS
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1894)
  AUTHORS   Oettgen,P., Finger,E., Sun,Z., Akbarali,Y., Thamrongsak,U.,
```

```
            Boltax,J., Grall,F., Dube,A., Weiss,A., Brown,L., Quinn,G., Kas,K.,
            Endress,G., Kunsch,C. and Libermann,T.A.
  TITLE     PDEF, a novel prostate epithelium specific ets transcription
            factor, interacts with the androgen receptor and activates
            prostate specific antigen gene expression
  JOURNAL   J. Biol. Chem. 275 (2), 1216-1225 (2000)
  MEDLINE   20092894
REFERENCE   2  (bases 1 to 1894)
  AUTHORS   Libermann,T.A., Oettgen,P., Kunsch,C. and Finger,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUN-1998) Medicine, Beth Israel Deaconess Medical
            Center, 330 Brookline Ave., Boston, MA 02215, USA
FEATURES             Location/Qualifiers
     source          1..1894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6p21.3"
                     /tissue_type="prostate"
     gene            1..1894
                     /gene="PDEF"
     CDS             416..1423
                     /gene="PDEF"
                     /note="prostate epithelium specific"
                     /codon_start=1
                     /product="Ets transcription factor PDEF"
                     /protein_id="AAC95296.1"
                     /db_xref="GI:4007418"
                     /translation="MGSASPGLSSVSPSHLLLPPDTVSRTGLEKAAAGAVGLERRDWS
                     PSPPAIPEQGLSAEYLSYFDMLYPEDSSWAAKAPGASSREEPPEEPEQCPVTDSQAPA
                     GSLDLVPGGLLTEEHSLEQVQSMVVGEVLKDTEIACKLLNITADPMDWSPSNVQKWLL
                     WEIEHQYRLPPMGKAPQELAGKELCAMSEEQFRQPSPLGGDVCHAHLLDWFSAAWMEER
                     TSPGATHYGASISEESWITGPVESCSGQGTHIWQFLKEILLKPDSYGKETEWLNKEK
                     GIFKIEDSAQVAPLWGTPENPPAMNYEFSRSTPQYYEKGILEKEIDSQRLVYQFVHP
                     I"
BASE COUNT       368 a    654 c    571 g    302 t
ORIGIN


alignment scores:
            Quality:     9.00              Length:       9
              Ratio:    1.000               Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-9458-2-COPY_66_74 x AF071538  ..

 Align seq 1/1  to: AF071538  from: 1  to: 1894

        1 LeuTyrProGluAspSerSerTrpAla 9
          ||||||||||||||||||||||||||
      611 CTGTACCCTGAGGACAGCAGCTGGGCA 637

seq name: qb_pri:AB031549

seq documentation block:
LOCUS       AB031549      1898 bp    mRNA            PRI       20-JAN-2000
DEFINITION  Homo sapiens PSE mRNA for prostate ets, complete cds.
ACCESSION   AB031549
VERSION     AB031549.1  GI:6721497
KEYWORDS    prostate ets.
SOURCE      Homo sapiens adult male prostate epithelium prostate cancer
            cell line:PC-3 cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Yamada,N., Tamai,Y., Miyamoto,H. and Nozaki,M.
  TITLE     Cloning and expression of the mouse Pse gene encoding a novel Ets
            family member
  JOURNAL   Gene 241 (2), 267-274 (2000)
  MEDLINE   20147594
REFERENCE   2  (bases 1 to 1898)
  AUTHORS   Nozaki,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-AUG-1999) to the DDBJ/EMBL/GenBank databases. Masami
            Nozaki, Osaka University, Institute for Microbial Diseases, 3-1,
            Yamadaoka, Suita, Osaka 565-0871, Japan
            (E-mail:mnozaki@biken.osaka-u.ac.jp, Tel:81-6-879-8348,
            Fax:81-6-879-8349)
FEATURES             Location/Qualifiers
     source          1..1898
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="PC-3"
                     /cell_type="prostate cancer"
                     /chromosome="6"
                     /dev_stage="adult"
                     /map="6p21.2-21.3"
                     /sex="male"
                     /tissue_type="prostate epithelium"
                     /note="vector:pUC18"
     gene            406..1413
                     /gene="PSE"
     CDS             406..1413
                     /gene="PSE"
                     /codon_start=1
                     /product="prostate ets"
                     /protein_id="BAA89544.1"
                     /db_xref="GI:6221498"
                     /translation="MGSASPGLSSVSPSHLLLPPDTVSRTGLEKAAAGAVGLERRDWS
                     PSPPAIPEQGLSAEYLSYFDMLYPEDSSWAAKAPGASSREEPPEEPEQCPVTDSQAPA
                     GSLDLVPGGLLTEEHSLEQVQSMVVGEVLKDTEIACKLLNITADPMDWSPSNVQKWLL
                     WEIEHQYRLPPMGKAPQELAGKELCAMSEEQFRQPSPLGGDVCHAHLLDWFSAAWMEER
                     TSPGATHYGASISEESWITGPVESCSGQGTHIWQFLKEILLKPDSYGKETEWLNKEK
                     GIFKIEDSAQVAPLWGTPENPPAMNYEFSRSTPQYYEKGILEKEIDSQRLVYQFVHP
                     I"
BASE COUNT       368 a    654 c    571 g    302 t
ORIGIN


alignment scores:
            Quality:     9.00              Length:       9
              Ratio:    1.000               Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-9458-2-COPY_66_74 x AB031549  ..

 Align seq 1/1  to: AB031549  from: 1  to: 1898

        1 LeuTyrProGluAspSerSerTrpAla 9
          ||||||||||||||||||||||||||
      601 CTGTACCCTGAGGACAGCAGCTGGGCA 627

seq name: qb_pat1:AX048159

seq documentation block:
LOCUS       AX048159      3317 bp    DNA                PAT       15-DEC-2000
DEFINITION  Sequence 12 from Patent WO0070092.
ACCESSION   AX048159
VERSION     AX048159.1  GI:11876975
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3317)
  AUTHORS   Kaufmann,J. and Xin,H.
  TITLE     Expression of ets-domain proteins in cancer
  JOURNAL   Patent: WO 0070092-A 12 23-NOV-2000;
            Chiron Corporation (US)
FEATURES             Location/Qualifiers
     source          1..3317
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

```
                     /note "GSEF-encoding sequence with promoter (Figs. 2A 2B)"
     promoter        1. .1480
                     /note "Promoter"
     CDS             1815. .2822
                     /note "GSEF coding sequence"
                     /codon_start 1
                     /protein_id="CAC19004.1"
                     /db_xref="GI:11876976"
                     /translation "MGSASGCLSSVSESHHLLPRGFVSSKLSLEEAAAACAVCLEEGEWV
                     PSPPATPEQCLSAFYLSYEDMLYPFDSSWAAKAPQASSEEPPPEEPEQCPVIDSQAPA
                     GSLDLVPGGLTLEEHSLEQVQSMVVGEVLKDLETACKLLNLIADLMDWSPSNVSKWLL
                     WTEHQYKLPPMGKAPQETACKFLCAMSFFQFFQPSPLGCDVLHARLDLWKSAAWMEEP
                     TSPGATRYCASLSEPSWTLSFVDSSCSGSEIHLWSFLFELLSPHRSYCEEPWLNEEE
                     GIFKIEDSAQVARLWGIKFNFPAMNYLKFSKSPQYYFFCSLTEEPLTSQPLVVSFVHE
                     I"
BASE COUNT       710 a    1026 c     970 g     611 t
ORIGIN


alignment_scores:
             Quality:     9.00              Length:       9
               Ratio:    1.000               Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_66_74 x AX048159  ..

 Align seq 1/1  to: AX048159  from: 1  to: 3317

        1 LeuTyrProGluAspSerSerTrpAla 9
          ||||||||||||||||||||||||||
     2010 CTGTACCCTGAGGACAGCAGCTGGGCA 2036

seq_name: gb_pr6:AL157472

seq_documentation_block:
LOCUS        AL157472   150660 bp    DNA             PRI       29-NOV-2000
DEFINITION   Human DNA sequence from clone RP11-375E1 on chromosome 6, complete
             sequence.
ACCESSION    AL157472
VERSION      AL157472.18  GI:11434622
KEYWORDS     HTG.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 150660)
  AUTHORS    Skuce,C.
  TITLE      Direct Submission
  JOURNAL    Submitted (29-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
             CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
             requests: clonerequest@sanger.ac.uk
COMMENT      On Nov 30, 2000 this sequence version replaced gi:11342772.
             During sequence assembly data is compared from overlapping clones.
             Where differences are found these are annotated as variations
             together with a note of the overlapping clone name. Note that the
             variation annotation may not be found in the sequence submission
             corresponding to the overlapping clone, as we submit sequences with
             only a small overlap as described above.
             This sequence has been finished according to sequence map criteria
             as follows:  An attempt is made to resolve all sequencing problems,
             such as compressions and repeats, but not necessarily within known
             annotated repeat sequence elements.  Where the sequence is
             ambiguous, there is an annotation using the 'unsure' feature key.
             The following abbreviations are used to associate primary accession
             numbers given in the feature table with their source databases:
             Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
             on the WORMPEP database can be found at
             http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
             was generated from part of bacterial clone contigs of human
             chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
             Group.  Further information can be found at
             http://www.sanger.ac.uk/Pub/6/F
             RP11-375E1 is from the library RPCI-11.2 constructed at the Roswell
             Park Cancer Institute by the group of Pieter de Jong. For further
             details see http://bacpac.med.buffalo.edu/
             VECTOR: pBACe3.6
             IMPORTANT: This sequence is not the entire insert of clone
             RP11-375E1 it may be shorter because we sequence overlapping
             sections only once, except for a 100 base overlap.
             The true left end of clone RP11-375E1 is at 1 in this sequence. The
             true left end of clone RP3-392O22 is at 150561 in this sequence.
             The true right end of clone RP11-302L10 is at 113044 in this
             sequence.
FEATURES             Location/Qualifiers
     source          1. .150660
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /chromosome="6"
                     /clone="RP11-375E1"
                     /clone_lib "RPCI-11.2"
BASE COUNT     45471 a   38887 c   38740 g   47563 t
ORIGIN


alignment_scores:
             Quality:     9.00              Length:       9
               Ratio:    1.000               Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_66_74 x AL157472/rev  ..

 Align seq 1/1  to reverse of: AL157472  from: 1  to: 150660

        1 LeuTyrProGluAspSerSerTrpAla 9
          |||| ||||||||||||||||||||||||
  129545 CTGTACCCTGAGGACAGCAGCTGGGCA 129519

seq_name: gb_pi3:GMCHLH

seq_documentation_block:
LOCUS        GMCHLH        4449 bp    mRNA            PLN       15-APR-1998
DEFINITION   Glycine max mRNA for magnesium chelatase subunit.
ACCESSION    AJ001094
VERSION      AJ001094.1  GI:3059094
KEYWORDS     chlH gene; magnesium chelatase subunit.
SOURCE       soybean.
  ORGANISM   Glycine max
             Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
             Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
             Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE    1  (bases 1 to 4449)
  AUTHORS    Nakayama,M.
  TITLE      Direct Submission
  JOURNAL    Submitted (13 AUG 1997) Nakayama M., Bioscience and Biotechnology,
             Tokyo Institute of Technology, Nagatsuta 4259, Midori ku, Yokohama,
             Kanagawa 226, JAPAN
REFERENCE    2  (bases 1 to 4449)
  AUTHORS    Nakayama,M., Masuda,T., Bando,T., Ohta,H. and Takamiya,K.i.
  TITLE      Cloning and characterization of the soybean chlH gene coding for a
             subunit of magnesium chelatase
  JOURNAL    Unpublished
FEATURES             Location/Qualifiers
     source          1. .4449
                     /organism "Glycine max"
                     /strain "Resnik"
                     /db_xref "taxon:3847"
                     /clone_lib="Clontech"
     gene            132. .4283
                     /gene "chlH"
     CDS             132. .4283
                     /gene "chlH"
                     /codon_start 1
                     /product "magnesium chelatase subunit"
                     /protein_id "BAA15251.1"
```

```
                     /db_xref "GI:305909"
                     /db_xref "SPTREMBL:O65808"
                     /translation "MASLVSGFTLGLLSAFLLLLLAGNEAHAYNCGKK
                     SSLRVKCAVIGNGLFTQTIQEVRSIVFENNQNLTIVRTVYVVLLAAYGSSIHAAVTAL
                     NSKKKHIASEFVVVGYLVEFFMDAATYKIFCKHIFDANIFIGSLIEVFFLAFFKAAVDK
                     ERDPLDAVIVFFSMPFVMPLNKFCSFSMSLGVQSKSPEFQGVAKKKPQGVACGAQSMCK
                     LVRTLPKVLKYLPSDKAGFAKLYISTQGWLCSNIGNLQNLFMLCSGYIPALKGTKET
                     EYSEPVLYLDVGTWHPLAPCMYDDVEEYLNWYCIKEPANLKLKSPNALVFSLVFLQPS
                     IVTGDDCHYVAVIMELEARCAKVTPIEACGLDESCPVEKFFETLELKKFFVNSVVSLI
                     GFALVGGPARQDHPRAVEALMKLDVPYTVALPLVPQTTEEWLNSTLGLHPTQVALQVA
                     LPELDGGMEPTVEAGRDFKTGKSHALHKRVEQLCTRATRWAELKRKSSAEKKLATTVF
                     SFPPPDRCNVCIAAYLNVEASEYSVMKELKKFGYNVPCLPETSPALTEDVILHDKEAQES
                     SPNLNTAYKMNVKELYQNLTTYATALELNWCKFPCNLNACCLNLLVYCGQYGNVFTGVQ
                     PTFGYEGDPMRLLFSKSASPHHGFAAYYSFVEKTFKADAVLHPGTRGSLKRMPCGKQVG
                     MSDVCYPDSLIGNTPNVYYYAANNPSEATTAKRRGYANTLSYLTPPAENAGLYKGLKQ
                     LSELISSYQNSLKTGGMCAQIVSSTLSTAKQWNFDKTVTTFQECKFTPPKFRTLVVCQV
                     YSKIMKTESRLLPCGLHTIGKPPSALEAVATLVNTAALLRPFDGISSLPSTLADTVGR
                     DTEDVYRGSNKGTLKDVELLRQTLRASGATLAPVERTLNNKGQVVDVADKLSSTTSF
                     GTNEPWTLYLCNTKTYAQCFTKTLTVTFTCCTLPSTVAGNLVCLPQGADTGPYVLCG
                     PGGDPTRNPKVLACTGKNTHALIDPQSTFTTAAMQSAKTVVDRLTRKQKAENGGKYTFTL
                     ALVLWGTDNTKTYGESLAQVLWMTGVEPVADTFGRVNKVEPVSLEELGRPETDVVVNC
                     SGVFRDLFTNQMNLLDPAVKMVAELDEPAEQNYVKKHASPQAQAIGVEVREAATRTFS
                     NASSSYSSNINLNIAVPNCSWNGKFEQLTMYLCWSGACTGAFTAMDLKPFVTLMAIC
                     TADATPQNLDSSFTSLTDVSHYFDSDPTNLVQNLRKQCKKPSAYIADFTTANAQVRTT
                     SFTVRTLDAFCKTLNLKWYFGMLCTPYTEVVSLTEFRTLTVSWSATLCQVFNWVYTLAN
                     TTFTQDEQMLNKLMSTNFNGLDFKKLVQFTLLANCGCYWLTSLFNTLKDLTLVSLVLKG
                     EGIDR"
BASE COUNT      1259 a     885 c    1074 g    1231 t
ORIGIN


alignment scores:
            Quality:     7.00            Length:       7
              Ratio:    1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_66_74 x GM*CHIB/rev

 Align seq 1/1   to reverse of GM*CHIB from 1 to 4427

        1 LeuTyrProGluAspSerSer 7
          |||||||||||||||||||||
     4427 CTCTATCCAGAAGATTCATCT 4427

seq_name: qb in3:DDB20432

seq_documentation_block:
LOCUS       DDB20432      6115 bp    DNA             INV       11-FEB-1995
DEFINITION  Dictyostelium discoideum TaqB (taqB) gene, complete cds.
ACCESSION   B20432
VERSION     B20432.1  GI:664839
KEYWORDS    .
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
REFERENCE   1  (bases 1 to 6115)
  AUTHORS   Shaulsky,G., Kuspa,A. and Loomis,W.F.
  TITLE     An MDR transporter/serine protease gene is required for prestalk
            specialization in Dictyostelium
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 6115)
  AUTHORS   Shaulsky,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JAN-1995) Gad Shaulsky, Department of Biology 0322,
            University of California, San Diego, 9500 Gilman Dr., La Jolla, CA
            92093, USA
FEATURES             Location/Qualifiers
     source          1..6115
                     /organism "Dictyostelium discoideum"
                     /strain "AX4"
                     /db_xref "taxon:44689"
                     /clone_lib "Lambda Zap II/4% (Shaulsky and Loomis,
                     Department of Biology 0322, Univ. of California, San
                     Diego)"
                     /clone_comm "4"
                     /map "5.2 Mb from ERS end"
                     /cell_type "prestalk"
                     /dev_stage "slug"
     misc_feature    1..6115
                     /phenotype "developmental arrest at tight aggregate stage.
                     Fst A-cell development missing."
     mRNA            <1..>5883
                     /gene "taqB"
                     /product "TaqB"
     gene            1..>5886
                     /gene "taqB"
     5'UTR           <1..168
                     /gene "taqB"
     CDS             169..>5886
                     /gene "taqB"
                     /note "serine protease/MDR transporter homolog, prestalk
                     specific"
                     /codon_start=1
                     /product "TaqB"
                     /protein_id "AAA62212.1"
                     /db_xref "GI:664840"
                     /translation "MKFQLSSSFLFLLSAVFLLGFSLPFLEGSSLSNGNENIFKFNN
                     INEFNYNGTKSTYKNKQQQQQFSNKTYGNLKKTLKNFLNPLTFIKFSINTNNNNNN
                     NNKLNNNNNNNNNNNNNNNNNNNNNNNNNNNYYNSTEYYSSFVSKFLKSNDDDGTYY
                     DDYQSKYFKNHYTVQKKDPRNDETKFQLKFFLTGTDTTTLKFQEFKSHIVHYTPHDSF
                     LVPMTKEQSVLLSSKEWTSWTGEHEPSNKTHLNYHKKSTGYPVYTTLSGSTNSLTQRW
                     ENTLNSTLTSYNSKVKLTLTNQKKLKSTVYCNDESFSSSSSSSCSLTGSEKTVYKWTS
                     EQSESNYTERSEKLQPANRLSPTVTFGTKDKLVNNDRTDTFPLRGKGQTTLSTADTGLDG
                     SHCFFESDSKYPTPFNQVNENHRKVVTYTTYHDNEDYVNGHGTHVCGSAAGTPEDSSWA
                     TSSFSGLATDAKIAFYDLSSGSSEPTPPEDYSQMYKPLYDAGAKVHGDSWGSVSLQGY
                     YGGYSDDAGTIDAFLYEYPEFSTLRAAGNNELFASLLAQATAKNATTVGAEQTAHVNY
                     VSDALEYYDFSDNANFQRPCLFDKKYCNYTTAKCCSEVSNVKGTQLQCCPASTKQNASD
                     SFTTQPQFYNENNMGSFSSKGPTHDGKLKPDTVAPGEYTTSARSNGENSTDQCGDGSL
                     PNANGLMSTSGTSMATPLATAATTTLRQYLVDGYFPTGESVEENKLLPTGSTLKALMI
                     NNAQTTNGTYTWSASSTNPSNATTFQTNGANTTQCWCAIFMNNWTYVFSSNPTPPSRW
                     TGTGGLGKNQKATEWKEDSLSSGLNKSYCFEYKPSSSSSGSGGGGGTPKTVAILVWTD
                     PDSYSGAKFNLVNNLDLLLLNSDDDSTTTGNSGGSTQFAGKVAQPFTLNNVFGTTIN
                     PIKAMNYKFTTAGTNVPTGPQKFSFVFHGENGQFDWADSCPQCVDGVQFPCLTTNGTG
                     TQSCGSDLLWTKCTVQSCNEGYNYNSTKNTCDKFTSYNYTTTTVAGGTMVLTTLLLMW
                     IKYQEYKESKKDSFERRFDFGTGTFVKREKDKDAKVTPPDLYNEVSPTTELLTAIACSL
                     VAIAASTQPFYIGNIVNNTCTTKSTGEFKSDFTTTFTLAFIEFLFINVGSWTSCIVN
                     EKMVMRIQNKVFKALTAQDMGFFQKNNSALLMNVLIVDTPMLRSALIGTLLSTATGVC
                     ELVCSLVETFTTSWELSLAFFAAVPTTQLVDQTQSQFTKSTLTPQTLFHNSKASQHGTE
                     SLTNMEVVTNYCKGEKEMVKYSIQLMNVFTTAEKLETTNTEAGTGKWLTTESLTFVTL
                     YFSAYTVTQKQFTVGTMTSFSLYTGYVVDASSSTFCVYVSYTQCTLASATKVFPMTLRSA
                     PRRPTTLEEEFEAPRLACLSCCCCCCCNCNCNFEEDFQENTENCFDVLTSNTTTTTDNVFN
                     SNCKQQDDNNNNNNTGNLDYSEQLTGVSTVADSTVGTTKPETKKQKEEEFQKEYYPKQTG
                     TSVAETPTFTLPSSYTELTECRGETEFKNVSFRYTPRPDVQVLHNTNMKFFAGKCYGLV
                     GPSGSGKSTTLLLTSKEYPLHGEIGGKTYMDGTDIAKTRPNNLRSFVTNVHQHPFTFD
                     ATTCENTCYATDNPTQEDTTEAAKLAYAHEFINDLPKKYDTQTGEAGNLSGGQKKRIA
                     VAKATCAKRKTMLLDEITAALDPESEEAITDSTKVTQGHTVVMVAHKVAAVRDCDKT
                     FVLEEGYLVEEGTHEELMANKGKYYFMFSEDKDPTPTLQNNNNNKNNNNNNNNEPSSS
                     SIPPNDQPTPPPQEQQEQKNDQPPPPTPQEQQEQQEQQQQQQQEQQQQQQQQQQQQQ
                     QQQQQQQQQQQQQQQQQNDQPPNDYDQVPPPPPLPSESPSPTPIGNNIGQPLVQMDEE
                     NDEER"
     misc_feature    1300..2266
                     /gene "taqB"
                     /note "encodes serine protease homology"
     misc_feature    3106..4368
                     /gene "taqB"
                     /note "encodes MDR transporter homology transmembrane
                     domain"
     misc_feature    4711..5238
                     /gene "taqB"
                     /note "encodes MDR transporter homology ATP binding
                     domain"
BASE COUNT      2376 a     844 c     926 g    1969 t
ORIGIN
```

```
alignment_scores:
            Quality:    7.00            Length:       7
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_66_74 x DDBJ:D20432   ..

 Align seq 1/1   to: DDBJ:D20432   from: 1   to: 6115

        3 ProGluAspSerSerTrpAla 9
          |||||||||||||||||||
     1498 CCAGAGGATTCTTCATGGGCT 1518

seq_name: gb_ba3:SC7H2

seq_documentation_block:
LOCUS       SC7H2          42655 bp    DNA             BCT       02 AUG 1999
DEFINITION  Streptomyces coelicolor cosmid 7H2
ACCESSION   AL109732
VERSION     AL109732.1  GI:5689879
KEYWORDS    amino acid decarboxylase; aminopeptidase, ATP binding; RNA helicase;
            formate dehydrogenase, iron-sulphur subunit; monooxygenase; mutase;
            periplasmic branched chain amino acid binding; permease; polA; DNA
            polymerase I; pyk; pyruvate kinase; response regulator; rpsA; 30S
            ribosomal protein S1; secreted nucleotidase.
SOURCE      Streptomyces coelicolor A3(2).
  ORGANISM  Streptomyces coelicolor A3(2)
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 42655)
  AUTHORS   Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J.,
            Kinashi,H. and Hopwood,D.A.
  TITLE     A set of ordered cosmids and a detailed genetic and physical map
            for the 8 Mb Streptomyces coelicolor A3(2) chromosome
  JOURNAL   Mol. Microbiol. 21 (1), 77-96 (1996)
  MEDLINE   97000851
REFERENCE   2  (bases 1 to 42655)
  AUTHORS   Saunders,D.C. and Harris,D.
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 42655)
  AUTHORS   Bentley,S.D., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-1999) Streptomyces coelicolor sequencing project,
            Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
            David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
            Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT     Notes:
            Streptomyces coelicolor sequencing at The Sanger Centre is funded
            by the BBSRC and Beowulf Genomics
            Details of S. coelicolor sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL: http://www.sanger.ac.uk/Projects/S_coelicolor/)
            CDS are numbered using the following system eg SC7H7.01c, SC (S.
            coelicolor), 7H7 (cosmid name), .01 (first CDS), c (complementary
            strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous.
            The length in codons is given for each CDS.
            Usually the highest scoring match found by fasta 3.0 is given for
            CDS which show significant similarity to other CDS in the database.
            The position of possible ribosome binding site sequences are given
            where these have been used to deduce the initiation codon.
            Gene prediction is based on positional base preference in codons
            using a specially developed Hidden Markov Model (Krogh et al.,
            Nucleic Acids Research, 22(22) 4768-4778(1994)) and the 'Frame1.0'
            program of Bibb et al., Gene 30:157-66(1984) as implemented at
            http://www.nih.go.jp/
            jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the
            correct initiation codon. Where possible we choose an initiation
            codon (atg, gtg, ttg or (att)) which is preceded by an upstream
            ribosome binding site sequence (optimally 5-13bp before the
            initiation codon). If this cannot be identified we choose the most
            upstream initiation codon.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
            sequenced clone. It may be shorter because we only sequence
            overlapping sections once, or longer, because we arrange for a
            small overlap between neighbouring submissions.
            Cosmid 7H2 lies between and overlaps with cosmids 8A3 and 2C9 on
            the AseI-C genomic restriction fragment.
FEATURES             Location/Qualifiers
     source          1..42655
                     /organism="Streptomyces coelicolor A3(2)"
                     /strain="A3(2)"
                     /db_xref="taxon:100226"
                     /clone="cosmid 7H2"
     gene            1..273
                     /gene="SC7H2.01"
     CDS             <1..273
                     /gene="SC7H2.01"
                     /note="SC7H2.01, partial CDS, unknown, len: 91aa; similar
                     to the C-terminal region of several hypothetical proteins
                     both prokaryotic and eukaryotic eg. SW:Y704_SYNY3
                     hypothetical protein from Synechocystis sp. strain PCC6803
                     (150 aa) fasta scores; opt: 100, z-score: 147.5, E(): 0,
                     0.84, (44.9% identity in 88 aa overlap) and SW:UBI4_CAPHI
                     from Capra hircus (goat) (137 aa) fasta scores; opt: 96,
                     z-score: 140.4, E(): 1.7, (34.6% identity in 81 aa
                     overlap)."
                     /codon_start=1
                     /transl_table=11
                     /label=SC7H2.01
                     /product="hypothetical protein"
                     /protein_id="CAB52043.1"
                     /db_xref="GI:5689880"
                     /translation="DFVRQAEQVLANLTLQLKAVQSGLFHVLSTDVYVVSALTAALSA
                     VWDVVPASDLSIGPHSSTLLGVSCLGYPGQLVETTATAVVPESGRP"
     gene            270..821
                     /gene="SC7H2.02"
     CDS             270..821
                     /gene="SC7H2.02"
                     /note="SC7H2.02, possible acetyltransferase, len: 184aa;
                     similar to SW:YIAC_ECOLI hypothetical protein from
                     Escherichia coli (146 aa) fasta scores: opt: 178, z-score:
                     227.9, E(): 2.3e-05, (27.2% identity in 147 aa overlap).
                     Also has region of similarity to several
                     acetyltransferases eg. TR:Q08414 (EMBL:X73149)
                     nourseothricin acetyltransferase from Streptomyces noursei
                     (189 aa) fasta scores: opt: 115, z-score: 150.4, E():
                     0.47, (40.2% identity in 82 aa overlap). Contains Pfam
                     match to entry PF00583 Acetyltransf, Acetyltransferase
                     (GNAT) family."
                     /codon_start=1
                     /transl_table=11
                     /label=SC7H2.02
                     /product="hypothetical protein"
                     /protein_id="CAB52044.1"
                     /db_xref="GI:5689881"
                     /translation="MNKEAGARAGADVQAGPSDDAGAGVSIDAGAFAVVLPRARAVPA
                     RAAADVWLRSFAAALPIVVRPRSDSEVRAYFRHVVVERYEIWVAQEDGAVAGVMVLE
                     GEELSQLYLAPERRGCGIGDRFVALAKERRALGLSLWTFQVNAPARKFYERHGFRAVF
                     WTDGDRNEEREPDVRYEWRPRSA"
     misc_feature    475..740
                     /gene="SC7H2.02"
                     /note="Pfam match to entry PF00583 Acetyltransf,
                     Acetyltransferase (GNAT) family, score 43.70, E-value
                     4.2e-09."
     stem_loop       820..900
                     /note="Inverted repeat (Score 57, 23/24 (888) matches, 0
                     gaps) with potential 28 base loop."
     gene            complement(907..1878)
                     /gene="SC7H2.03c"
     CDS             complement(907..1878)
                     /gene="SC7H2.03c"
                     /note="SC7H2.03c, probable aminopeptidase, len: 323aa;
```

```
                     similar to many proline aminopeptidases eg. SW:PIP_XANCI
                     proline iminopeptidase from Xanthomonas campestris pv.
                     citri (313 aa) fasta scores: opt: 848, z score: 979.4,
                     E(): 0, (44.4% identity in 320 aa overlap). Contains Pfam
                     match to entry PF00561 abhydrolase, alpha/beta hydrolase
                     fold."
                     /codon_start=1
                     /transl_table=11
                     /label=SC7H2.04c
                     /product="putative aminopeptidase"
                     /protein_id="CAB52045.1"
                     /db_xref="GI:5689882"
                     /translation="MSLYPETEPYDSGMLDVGDGNHVYWETCGNPDGKPAVVLHGGPG
                     SRASPGLRRYFDPAAYRIVLLDQRGAGRSLPRASAPETDMSVNTTAHLMADIERLRVH
                     LGIERWLVWGVSWGSVLGLRYAQTHPGVVIELVLTGVATGSNAEVALLTRGLGNIFPE
                     AHERFLAELPTDARDGNLPAAYNRLLESPDPAVRERAARAWTDWETAIIPAPDGSVAR
                     YQDPDFRMGFARTVTHYWGNDHFLGDGNDEGVVIRDAHLLKGIPGTLVQGSLDFGNLL
                     GIVWRLHHAWPDSDLVIVDEAGHDAGTPGDEALLAAIDKYARGGIAE"
misc_feature         complement(928..1701)
                     /gene="SC7H2.04c"
                     /note="Pfam match to entry PF00561 abhydrolase, alpha/beta
                     hydrolase fold, score 74.20, E value 2.7e-16."
gene                 complement(1933..2646)
                     /gene="SC7H2.04c"
CDS                  complement(1933..2646)
                     /gene="SC7H2.04c"
                     /note="SC7H2.04c, unknown, len: 247aa; similar to
                     SW:YO8N_MYCTU hypothetical protein from Mycobacterium
                     tuberculosis (250 aa) fasta scores; opt: 870, z score:
                     990.4, E(): 0, (57.1% identity in 231 aa overlap)."
                     /codon_start=1
                     /transl_table=11
                     /label=SC7H2.04c
                     /product="hypothetical protein"
                     /protein_id="CAB52046.1"
                     /db_xref="GI:5689883"
                     /translation="MDTSCDVSSLKTCSKACTSCVPVLPTVARTESAAFAHWTVWS
                     RPVPGFGPPDARLLIVGLAPAAHGNEFGRMFTGDESGDVLYQALYDVGLASQPTAVR
                     VDDGLELYGVRVTSPVHCAPPANKPTPAERDICRSWLVQELGLLRPTLRAVVVLGAFG
                     WQAALPAFAGAGWTVPRPRPAFAHGTQVTLDAADGPDLHLFGCFHVSQRNTFTGRLTP
                     EMLRDVLRTAAETAGLPAR"
gene                 2763..3323
                     /gene="SC7H2.05"
CDS                  2763..3323
                     /gene="SC7H2.05"
                     /note="SC7H2.05, unknown, len: 186aa; similar to
                     SW:YREH_ECOLI hypothetical protein from Escherichia coli
                     (183 aa) fasta scores; opt: 228, z score: 249.5, E():
                     1.4e-06, (35.2% identity in 122 aa overlap)."
                     /codon_start=1
                     /transl_table=11
                     /label=SC7H2.05
                     /product="hypothetical protein"
                     /protein_id="CAB52047.1"
                     /db_xref="GI:5689884"
                     /translation="MASEHDEGGSGEVAAPKDGAAAAAGTPDPKTATPDPKTAAATAA
                     AEAAGPQHGESVRVDSWIWAVRLIKTRSLGATACRGGHVRVNGERVKPAYSLRVGDEV
                     RLRQEGRERVVVVKRLIRKRVGAPVAVQCYVDNSPPPPPREAVAPAGTRDRGAGRPTK
                     RDRRREMDRLRGLEGLSSSRRDAGTRP"
gene                 complement(3375..3881)
                     /gene="SC7H2.06c"
CDS                  complement(3375..3881)
                     /gene="SC7H2.06c"
                     /note="SC7H2.06c, possible integral membrane protein, len:
                     168aa; contains possible membrane spanning hydrophobic
                     regions."
                     /codon_start=1
                     /transl_table=11
                     /label=SC7H2.06c
                     /product="putative integral membrane protein"
                     /protein_id="CAB52048.1"

alignment_scores:
                 quality:     7.00              Length:       7
                   Ratio:    1.000               Gaps:       0
  Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US 09 126 945B 2 COPY 66 74 x SC7H2/rev   ..

 Align Seq 1/..  to reverse of: SC7H2  from: 1  to: 42655

        4 ProGluAspSerSerTrpAla 9
          |||||||||||||||||||||
    36703 CCGGAGGACAGCTCATGGGCA 36683

Seq name: gb:htg21:AL357145

Seq documentation block:
LOCUS        AL357145    92080 bp    DNA              HTG       23-JAN-2001
DEFINITION   Homo sapiens chromosome 1 clone RP4-699P10 map p42.1-42.3, ***
             SEQUENCING IN PROGRESS ***, 12 unordered pieces.
ACCESSION    AL357145
VERSION      AL357145.3  GI:9797675
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 92080)
  AUTHORS    Davitt,R.
  TITLE      Direct Submission
  JOURNAL    Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
             CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
             requests: clonerequest@sanger.ac.uk
COMMENT      On Aug 12, 2000 this sequence version replaced gi:9214086.
                   -- -- Genome Center
             Center: Sanger Centre
             Center code: SC
             Web site: http://www.sanger.ac.uk
             Contact: humquery@sanger.ac.uk
                   -- ---- Project Information
             Center project name: dJ699P10
                   - - -- Summary Statistics
             Assembly program: XGAP4; version 4.5
             Sequencing vector: plasmid; L08752; 100% of reads
             Chemistry: dye terminator ET-amersham; 22% of reads Chemistry:
             Dye-terminator Big Dye; 77% of reads
             Consensus quality: 86851 bases at least Q40
             Consensus quality: 88611 bases at least Q30
             Consensus quality: 89693 bases at least Q20
             Insert size: 90980; sum-of-contigs
             Insert size: 116487; 11.4% error; agarose 1p
             quality coverage: 3.56x in Q20 bases; sum of contigs quality
             coverage: 2.94x in Q20 bases; agarose 1p
                   - -
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 12 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence
             * as soon as it is available and the accession number will
             * be preserved.
             *       1     2456: contig of 2456 bp in length
             *    2457  2556: gap of      100 bp
             *    2557  10002: contig of 7746 bp in length
             *   10003 10402: gap of      100 bp
             *   10403  24179: contig of 12777 bp in length
             *   24180 24279: gap of      100 bp
             *   24280  39494: contig of 16215 bp in length
             *   39495 39594: gap of      100 bp
             *   39595  45448: contig of 5754 bp in length
             *   45449 45448: gap of      100 bp
             *   45449  50407: contig of 4959 bp in length
             *   50408 50507: gap of      100 bp
```

```
            *     50508    55360: contig of 4853 bp in length
            *     55361 55460: gap of      100 bp
            *     55461    62881: contig of 7421 bp in length
            *     62882 62981: gap of      100 bp
            *     62982    70522: contig of 7541 bp in length
            *     70523 70622: gap of      100 bp
            *     70623    74312: contig of 3690 bp in length
            *     74313 74412: gap of      100 bp
            *     74413    79454: contig of 5042 bp in length
            *     79455 79554: gap of      100 bp
            *     79555    92080: contig of 12526 bp in length.
FEATURES             Location/Qualifiers
     source          1. .92080
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /chromosome="1"
                     /map "p32.1-32.3"
                     /clone="RP4-699P10"
                     /clone_lib="RPCI-4"
     misc_feature    1. .2456
                     /note="assembly_fragment:00840
                     fragment_chain:1
                     clone_end:17
                     vector_side:left"
     misc_feature    2557. .10302
                     /note="assembly_fragment:00572
                     fragment_chain:1"
     misc_feature    10403. .23179
                     /note="assembly_fragment:00813
                     fragment_chain:2"
     misc_feature    23280. .39494
                     /note="assembly_fragment:00835
                     fragment_chain:2"
     misc_feature    39595. .45348
                     /note="assembly_fragment:00737
                     fragment_chain:2"
     misc_feature    45449. .50407
                     /note="assembly_fragment:00434
                     fragment_chain:3"
     misc_feature    50508. .55360
                     /note="assembly_fragment:00497
                     fragment_chain:3"
     misc_feature    55461. .62881
                     /note="assembly_fragment:00659
                     fragment_chain:3"
     misc_feature    62982. .70522
                     /note="assembly_fragment:00414
                     fragment_chain:4"
     misc_feature    70623. .74312
                     /note="assembly_fragment:00800
                     fragment_chain:4"
     misc_feature    74413. .79454
                     /note="assembly_fragment:00249"
     misc_feature    79555. .92080
                     /note="assembly_fragment:00801"
BASE COUNT    26190 a  17695 c  18092 g  [illegible]
ORIGIN


alignment_scores:
            Quality:     7.00            Length:       7
              Ratio:    1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2-COPY_66_74 x AL357145/reverse

 Align seq 1/1  to reverse of AL357145  from 1  to 92080

        1 LeuTyrProGluAspSerSer 7
          |||||||||||||||||||||
     8819 CTATACCCAGAAGATAGCAGC 8799
```

```
seq_name: ab_ptr3:N17E5

seq_documentation block:
LOCUS        N17E5      130049 bp    DNA              PLN       08-FEB-2001
DEFINITION   Neurospora crassa DNA linkage group V Cosmid contig 17E5.
ACCESSION    AL513467
VERSION      AL513467.1  GI:12718264
KEYWORDS     .
SOURCE       Neurospora crassa.
  ORGANISM   Neurospora crassa
             Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
             Sordariales; Sordariaceae; Neurospora.
REFERENCE    1  (bases 1 to 130049)
  AUTHORS    Schulte,U., Arm,V., Bohnsel,J., Brandt,P., Fartmann,B.,
             Holland,K., Nyakatura,G., Mewes,H.W. and Mannhaupt,G.
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 130049)
  AUTHORS    German Neurospora genome,project.
  TITLE      Direct Submission
  JOURNAL    Submitted (07 FEB 2001) MIPS, at the Max Planck Institut fuer
             Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
             mannhaup@mips.biochem.mpg.de Project Coordinator: Ulrich Schulte,
             Institute of Biochemistry, Heinrich-Heine-University, D-40225
             Duesseldorf, Email: ulrich.schulte@uni-duesseldorf.de
COMMENT      This contig is an assembly of cosmid 17E5 from 1 to 45348, bAC
             10A6 from 45345 to 108703 and cosmid 11BH7 from 108704 to 130049,
             cosmid library pWRI5I6Xh, strain 74 OR-23-1A; cosmid clones and
             bAC clone 10A6 (strain OR74A) are available at the Fungal Genetic
             Stock Center, http://www.fgsc.net
             Sequencing was performed by MWG Biotech AG, Ebersberg, Germany,
             http://www.mwgdna.com
             Information on performance of analysis and a more detailed
             annotation of this entry and other sequences of linkage groups II
             and V can be viewed at:
             http://www.mips.biochem.mpg.de/proj/neurospora.
FEATURES             Location/Qualifiers
     source          1. .130049
                     /organism "Neurospora crassa"
                     /db_xref "taxon:5141"
                     /chromosome "2"
     misc_feature    1. .2158
                     /note "overlap to cosmid contig 93G11, EMBL:AL513443"
     exon            complement(8141. .8294)
                     /gene "17E5.10"
                     /number 1
     gene            complement(join(8141. .8294,8398. .8685,8762. .8998,
                     9097. .9200,9333. .9483,9563. .9721,9832. .10033,
                     10723. .10772))
                     /gene "17E5.10"
     gene            8141. .10772
                     /gene "17E5.10"
     CDS             complement(join(8141. .8294,8398. .8685,8762. .8998,
                     9097. .9200,9333. .9483,9563. .9721,9832. .10033,
                     10723. .10772))
                     /gene "17E5.10"
                     /note "weak similarity to 4-hydroxybenzoate (EC 2.7.1.8),
                     homo sapiens, SWISSPROT:KHK_HUMAN"
                     /codon_start 1
                     /product "hypothetical protein"
                     /protein_id "CAC28827.1"
                     /db_xref "GI:12718265"
                     /translation "MKHLIIVGACYLDIIIEIAVAVIGMIKIIWMEYSPSSVIIIEII
                     PKLACSVPQFPDFDSKLKAIAIQVKRQGNCPNIAEVIQQLLLSGAGIGSSTQNQPDS
                     AAGVVAVIPPPSQEQQQQQQQLKLHLVSILPDVGSPATRKIISSEFATPIAPQPSPSY
                     DSPSIIPESVISAASAIATSVDFSHCLYPKGHDFAASSYIIRSGALPSKIIVNYNDIE
                     MPIIEFMFIAIGCFVSLQHDHNVIASSFPTFGFTCWWHPFVPSAIPVIPMGLAPSLHA
                     FRDAPYSPCRIPFFTLQCIPYIFQVIPPSTISVPVFKPCKECIAFIAAFSCVVFYSKS
                     WAKRLSLFPACDELLVLVHDFKAWSPSGCVEPFIRQFPSTSFPFAACLPESVKAGAL
                     FPSPAGIGGGNNEQIAEVAEPDYLRCSIAETERSFIANNSAYKIFLVHIFFSVSAGIR
                     GSQKANSINIANKVNFSFTHSFIGAGHIFIAGIMYSIIMASFSAIATAIATAIKIRSW
                     IQSISQEEFEELIAFAVRLATQEVQMEGFNCIVEGSWADFFILASDSGIHVKQSRWI
                     AAEQKSIIMGWFRSACDDDDGQFPSPHV"
     intron          complement(8295. .8397)
```

```
                     /number 1
     exon            complement(8398. .8685)
                     /gene "17E5.10"
                     /number 2
     intron          complement(8686. .8761)
                     /number 2
     exon            complement(8762. .8838)
                     /gene "17E5.10"
                     /number 3
     intron          complement(8839. .9096)
                     /number 3
     exon            complement(9097. .9200)
                     /gene "17E5.10"
                     /number 4
     intron          complement(9201. .9432)
                     /number 4
     exon            complement(9433. .9483)
                     /gene "17E5.10"
                     /number 5
     intron          complement(9484. .9562)
                     /number 5
     exon            complement(9563. .9721)
                     /gene "17E5.10"
                     /number 6
     intron          complement(9722. .9831)
                     /number 6
     exon            complement(9832. .10033)
                     /gene "17E5.10"
                     /number 7
     intron          complement(10034. .10722)
                     /number 7
     exon            complement(10723. .10772)
                     /gene "17E5.10"
                     /number 8
     gene            14107. .14875
                     /gene "17E5.20"
     exon            14107. .14162
                     /gene "17E5.20"
                     /number 1
     CDS             join(14107. .14162,14243. .14117,14205. .14378,14442. .14875)
                     /gene "17E5.20"
                     /note "similarity to UTR1 protein (associated with ferric
                     reductase activity)   YER045c, Saccharomyces cerevisiae,
                     PIR:S46589"
                     /codon_start 1
                     /product "related to UTR1 (associated with ferric
                     reductase activity)"
                     /protein_id "CAC28829.1"
                     /db_xref "GI:12718266"
                     /translation "MPVFLFTSGDHSRPERNAFTSINSIETRSPITRGLEIDIARDAVQ
                     VSRENSQDQVLDPKGFSSTVGNIAANVRAKFCSMTNANSSDTSDAGQVDEKFWKSEIV
                     KAPNIDFEEKYSKLLIKKQLSEMAWGVRELSKRLSSVKLKFKIKTIFVLTKPYDQELI
                     PKTRALIRWLLDKKRDVRYIVYTDEALRCNEKLDAPGLIEETPKDYVESSFISEEAGN
                     DISERLRYWDEEICKARPQTADFVTTGGDGTVLFASWLPQRTVPPVLSFSLGSLGFL
                     TKFDFDHYHETLIAAFTEGVNVALKLRFAGTVMKNGPDCQDENDPSEEKFFGFKRDV
                     EDLIGEEKDDERTHRPDGIWCVLNEVVVDRGPNPTMSYIETFGDEEHFTSVNADGICV
                     STPTGSTAYNLAAGGSLCHPENPVMLVTAMCAHFLSFRPTLLPDTIVLRTGVPYDART
                     GSWASFCGREVIFLRRGDYVFLSASEYTFASVQDQCKSFFWIKSINAKLSWNTRQFQ
                     KSFT"
     intron          14163. .14242
                     /gene "17E5.20"
                     /number 1
     exon            14243. .14117
                     /gene "17E5.20"
                     /number 2
     intron          14118. .14204
                     /gene "17E5.20"
                     /number 2
     exon            14205. .14378
                     /gene "17E5.20"
                     /number 3
     intron          14379. .14441
                     /gene "17E5.20"
                     /number 3
     exon            14442. .14875
                     /gene "17E5.20"
                     /number 4
     gene            15406. .18231
                     /gene "17E5.40"
     exon            complement(15406. .18231)
                     /gene "17E5.40"
                     /number 1
     gene            complement(15406. .18231)
                     /gene "17E5.40"
     CDS             complement(15406. .18231)
                     /gene "17E5.40"
                     /note "similarity to hypothetical protein SPBC947.10c,
                     Schizosaccharomyces pombe, PIR:T40772"
                     /codon_start 1
                     /product "conserved hypothetical protein"
                     /protein_id "CAC28829.1"
                     /db_xref "GI:12718267"
                     /translation "MPSQNSDVRALLIVLILLWIFFSPDSSDPASLIEPEIAHAREAR
                     FQSALDVLNLIRWGDFAPDAPSSISPEPPSEIDGPAGEIGKGNEGKHIQGPQFLNLIG
                     FSQIDEKGYAWEDLGRFRNLCREWSKNAFLSAAIVCFLFLSWFFSMVSCIWKNAIGI
                     LHGTWVRRPGTVVRQAAEYNLSAIAPGATWPSSLGFWGRNLIGDIGKIELRIDEDEGD
                     GVYEEKVEGKTSRGANLARELQAVATIQDDASEASSWSLLLHGVHWPRQGAILLLISS
                     EFFSIEGLPHLAPGPNFFHPGQHLINFLCKLIQARKKSKYGFIGNFWGSRMPLDDS
                     FDALCFYVMVIQLQPLFLSRQYFSPQLLESYIEELERELRHPTGAFIGKIPDLRASMV
                     AWSPLCSEELESEGPPEEPSVEGEHLVGMKEAMEIYIAEEWLIAEAAVMLGQIWLEKE
                     QMELSNIPSTTGRVSFWTIGAMLFAIGMIFVTACAWSLNAIILLLPCLLVIFAAFISM
                     FIFGAFLSFVWKIQFPERRIFERFFQAAAAAAAAAAANKNASSSSAASTIPPTTQP
                     EPITIQPAPPPPPSQPRQPTIVPSDGDIDEAEQLNNLLSGASALPLFAIAKQFGLPAP
                     INEPRPTTPFSTTTGREVLAGLFLLFISAVATSWPSHTRSFYVNTLSFLYLSLWIPQI
                     IKNAQKNSKQAFSWKFMIGQSLLKLLFFAYFFLREDNVLLADPDFFAFLVLVAWVWIQ
                     LWVICAQSTLGPKFGVPKGWVKEAWDYHPVIRHDDLEGVSLPTGARKCDGLSSSSSS
                     SSPIEEKVRGKAWSVVSLGGPAAAAETGEGSAAAAAAAAAIDSAGQTRGNEEKEKEER
                     EKQKGVIMKSVDCAICKELLEVPVFNNNKGSGGSSDSAASGTTGVFARKAYMVIPCKH
                     IFHTNCLEGWMKYRLQCPICREELPPL"
     repeat_region   15718. .15741
                     /note "24 bp aqcaqcaqcaqc tandem repeat"
     repeat_region   16595. .16624
                     /note "30 bp qcaqcaqcaqcaqcaqca tandem repeat"
     exon            18843. .18897
                     /gene "17E5.40"
                     /number 1
     gene            18843. .19153
                     /gene "17E5.40"
     CDS             join(18843. .18897,18981. .19153)
                     /gene "17E5.40"
                     /note "weak similarity to ribonuclease precursor, Bacillus
                     pumilus, PIR:S51002"
                     /codon_start 1
                     /product="hypothetical protein"
                     /protein_id="CAC28830.1"
                     /db_xref "GI:12718268"
                     /translation "MSENGEANTSFSFIAINAKLSVQLEDGLNTLSNKIAGVGSEIFA
                     KSKSLDADYKLKSVKSVNANMLGQWTRCRGV"
     intron          18898. .18980
                     /gene "17E5.40"

alignment scores:
            quality:     7.00              Length:       7
              Ratio:  1.000                Gaps:         0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 RS 09 126 945B 2 COPY 66_74 x NC17E5  ..

 Align seq 1/1  to: NC17E5  from: 1  to: 140049

        8 ProGluAspSerSerIlpAla 9
          |||||||||||||||||||
   95611 CCCGAAGATTCTCCTGGGCG 95631

seq name: gb_pri:AC006423
```

```
Seq documentation block.
LOCUS       AC006423    123692 bp    DNA             SYN       07-SEP-2001
DEFINITION  Arabidopsis [illegible] BAC [illegible] genomic sequence,
            complete sequence.
ACCESSION   AC006423
VERSION     AC006423.7  GI:6939155
KEYWORDS    HTG.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 131692)
  AUTHORS   Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hanson,N.F.,
            Altafi,H., Nguyen,M., Tam,B., Schwab,S., Miranda,M., Preuss,D.,
            Buehler,E., Chao,Q., Chin,C., Chiou,J., Choi,E., Gonzalez,A.,
            Howung,B., Johnson-Hopson,C., Khan,S., Kim,C., Kim,T., Lee,J.M.,
            Lenz,C., Liu,A., Liu,S., Mukharsky,N., Pham,P., Sakano,H.,
            Shinn,P., Toriumi,M., Vaysberg,M., Yu,G., Ecker,J., Theologis,A
            and Davis,R.W.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 131692)
  AUTHORS   Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hanson,N.F.,
            Altafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
            Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
            Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V.,
            Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JAN-1999) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
REFERENCE   3  (bases 1 to 131692)
  AUTHORS   Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hanson,N.F.,
            Altafi,H., Nguyen,M., Tam,B., Schwab,S., Buehler,E.,
            Chin,C., Chiou,J., Choi,E., Dunn,P., Gonzalez,A., Howung,B., Kim,C.,
            Kim,T., Lee,J.M., Lenz,C., Liu,A., Liu,S., Luros,S.,
            Mukharsky,N., Pham,P., Sakano,H., Schwartz,J., Shinn,P.,
            Thaveri,A., Toriumi,M., Vaysberg,M., Walker,M., Yu,G., Ecker,J.,
            Theologis,A. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-FEB-2000) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
REFERENCE   4  (bases 1 to 131692)
  AUTHORS   Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hanson,N.F.,
            Altafi,H., Nguyen,M., Tam,M., Schwab,S., Preuss,D., Buehler,E.,
            and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-APR-2000) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
COMMENT     On Feb 8, 2000 this sequence version replaced gi:6688726.
FEATURES             Location/Qualifiers
     source          1..131692
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="1"
                     /clone="F28H19"
     gene            complement(4756..7438)
                     /gene="F28H19.1"
     CDS             complement(join(4756..[illegible],6[illegible]..7438))
                     /gene="F28H19.1"
                     /note="Hypothetical protein"
                     /codon_start=1
                     /protein_id="AAF63113.1"
                     /db_xref="GI:7523673"
                     /translation="MESVESSVEELKELWHLLPEEITPCSENESSYNIETWKQRVEGNG
                     KKESSSTIPELLIGNVSPNEIVTSTVSVPASQVFPSAWVPESIIGVAESFVK
                     PPESHLPESVLVSAIPSELSEFYTPSWGAMSSKSKRRWRHKWNLESSSAADHTSKILV
                     GDFDQIAATSDHYSVLQTSIPMRGLEEFQNCLNESIDLVDHFSRSYVHYTWSNHYDNI
                     IRKIDRAIANCEWPSSFPSATAVEFLSVSEHHSLSTILLENGFKESPETSFYESFLSS
                     HPTFLVSLIVAWEEQIPVGSHMPSLGFHLKAAKKQVKELNKVGSLNLQRLKSFALDSL
     gene            7655..9421
                     /gene="F28H19.2"
     CDS             join(7655..8355,8480..8589,9117..9421)
                     /gene="F28H19.2"
                     /note="Hypothetical protein"
                     /codon_start=1
                     /protein_id="AAF63112.1"
                     /db_xref="GI:7523672"
                     /translation="MIRTINNVIRALLFQARLSPCYWVFALNVVVHILKTLPSIAIQN
                     [illegible]
                     DLKTNKIITSKYVYFDEDTFPAINQNLSQVSSYTFLNAVIDPSPLFKTLQSELPLNQ
                     [illegible]
                     AMIKIKFRIWTHDLLRDFQSGSELKTIRVVLDFGVALNWFTQQIDVKNAILQFVSMVI
                     YSIKHSKDRIYHIQSIRFAYTCMNALERTICYLQTIKSYGLHLVKDNISILLIYSGAD
                     WDGCPDIQRSTLSGYCVFLGPNLVPWSAKRQPTVSRSS"
     gene            24882..26655
                     /gene="F28H19.3"
     CDS             join(24882..24899,24997..25082,25190..25906,[illegible]..[illegible],
                     26354..26481,26543..26655)
                     /gene="F28H19.3"
                     /note="Hypothetical protein"
                     /codon_start=1
                     /protein_id="AAF63102.1"
                     /db_xref="GI:7523662"
                     /translation="MSNLHVKGPVFQPGDIGFEEAIVFLFSCMESSIRSYGLQTTFIP
                     PIEYITWICEDGDNSESDSDCNEVDQTVKKKKKSRNQPLVVLAEDNVIQEPLQSSKRE
                     PSLYSRPFHFIGIGLVLLAADGFVFSGFVPGFGIAFEEIVQLIAMGCFRSSK
                     EEEEEEEEKSLNHSPFVLAVFIHEIQVLTNVVFDVFVSSSSPTKFIMESFRQFESSDSK
                     EDHELGVLGSELTANSSSVPEHNSCAVEEEELSSCNKQTQLAMSSSCEVALSNMPL
                     TLISRNYRAQFIKIPLWRGLMSVRGGNSCIMIGIVADVSSLACPAVHLASSLKGDLS
                     ALIGSNLEEVESLVIEMEKNDSAMPCVLNIDAELLLESYMLPSESWEEPVDYRIQL
                     LLVSTTGASSSLDRILLDISHSKIRYL"
     gene            complement(28643..32897)
                     /gene="F28H19.4"
     CDS             complement(join(28643..29382,29592..29649,[illegible]..[illegible],
                     31044..31190,31455..32291,32699..32897))
                     /gene="F28H19.4"
                     /note="Similar to gag-pol polyproteins"
                     /codon_start=1
                     /protein_id="AAF63111.1"
                     /db_xref="GI:7523671"
                     /translation="METPKPAILKQENYLIWEPELPEIVLSYEQIWKPVFTPEVGEKLP
                     NKTAVPVSTGKKFFPSSENDPSVQKPQVSLPLFPCPGPLLIAHGCIYSSEPKKAWVCD
                     WEEECEMEPCWITLHLKNAKRANGSCLVAVPSCLAVFPSNLEALTSSIASLKES
                     GISPTTYEPSKILKESDISFISEPSKMLVIHSGASHHMINNPSLIESIKLAIGNVVI
                     ANADKVVVKGIGNLNLDPKKSKALEMPSELILLESVPRAINLNCYTLPGPSSVHFQD
                     IEEPVILKESPAPGLFYVLEKIFLSPWIGVSNGNVAWHAPCVGPHARAFIL
                     MLVNVSFDNSIPIFIPQWWPVYFSQIQKVPKRIWHYSSNFISLQIPCRWSGPKKISIP
                     HGVSGAQKNKLIAKSIPCMEIGYSIIQKEYKCYEPINGPLYVSKIVKFLEPCYEKK
                     DWNSLADISTPSIDPAFSLQFLLDHLGVTPSSEPPFIKIRDLVEEIVVIIQQHPQEFIP
                     NLQQDGPVSGIPIYIDDVDILIQHEDVSAQELIDQHIIIGQSELFPGAISLIPPSIPIK
                     KKSCELYYGNEAVAHIIQAVCSLALLPCEPAFLCKIELNFVPCIYIEAKFSPEWMIA
                     VAPEIIGAMIKNHIWDPEELDQKKSALHTVRVVLSFAINISWDIWQVNVKNAFIQGFIE
                     EEVYMIPPPGLEDSIAIGKVFRLRKAIYGLKQSPRVWYHKISTILKEKCKKSEADHI
                     LIAIQSKAGIIVVLIYVDDILIIGDDNKLANKGKKKNNAIIDFLNIGYIELSQYKKLVG
                     [illegible]
                     IVGYCHADYGRDITDRKSTIGYCIFIGGNLVTWKTKKQKTVSCSSAESEYRAMKELIN
                     [illegible]
                     CMILSCYIDTKPNLIDIFIKGASRKVCEHIYNKLIGEVDLIEP"
     gene            40184..43772
                     /gene="F28H19.5"
     CDS             join(40184..40448,41154..41261,41361..41453,41537..41718,
                     42922..42989,43522..43710,43796..43844,43914..43929,
                     43605..43772)
                     /gene="F28H19.5"
                     /codon_start=1
                     /product="Putative serine carboxypeptidases"
```

```
                     /protein_id "AAF63101.1"
                     /db_xref "GI:7523661"
                     /translation "MVCPEWEELVAVVVMVLAWPVYNCNLAVCPVQEVKELPGL
                     PEVAFRQFAGYVDIDVKAGRSLFYYFVEAEKQPHSKPLTLWLNGGPGCSSIGGAFTE
                     LGPFYPTGDARGLRRNPKSWNKASNILFVDSPAGVGWSYSNTESDYTTGDFSTAKDML
                     VFMLRWLEKFPQFKTRNLFLAGESYAGHYVPQLADVILEYNAQRSNRPKFNLKGIAIG
                     NPLLKLDRDVPATYEFFWSHGMISDELGLEIMNQCDFEDYIFTDSHNISKLCAAVNQ
                     AGTITTQYVNYYDILLDVCYPSLEFQQELKKKMGIRMSFGVDVCMSEEEQLEYLNLEV
                     QKALHANKTKLPYEWSMCSSILNYKYTLCNANMLFLNKFVKSKVPVWVFSGIEDSVI
                     PLLGSRTLVKELADDLNFNTIVPYGAWFDKGQVGGWVVEYGNLLFFAILVRGAAHMVPY
                     SQPSRALHLFTSFVLGRKLPHKSDPALHD"
     gene            complement(47142..51283)
                     /gene "F28H19.6"
     CDS             complement(join(47142..47798,47864..48831,48916..49108,
                     49221..50279,50708..51283))
                     /gene "F28H19.6"
                     /codon_start 1
                     /product "putative retrotransposon pol polyprotein"
                     /protein_id "AAF63110.1"
                     /db_xref "GI:7523676"
                     /translation "MSEIVPAAPCPEEVSSSSSQCPMCNATNYIVWAMKMKRALEVH
                     KAWNMIEKVIELPSDKSCVACALLFLCIFLALALLTCLLTARLVWLAIFIHECVE
                     RVKDARLQTLMAEFDRLIMKISEKLYDFAGKLSETSAKSAALGVLTEESKLVKKELKS
                     VPRRKYIHIVAALEQVLDLNLCLIFQDIVGRLKKPYDRGYKVESTLDDSVSGKVHFRDD
                     SRIDIKGNGQQQKVSNVYITLELKSNFIGIGQATLSGCNTIMKGDQLTMHDRGGKLLV
                     KACRSKNHLYKVRMGTKGAIRHLITISESSRWHHARLGHVNFQILKSMMQKELVQGFS
                     LGSTKKELCSSCLLAGKQTRLAFPKATLERAAKTLELLVHGDLCGPITPSTAAGNKYVFV
                     LIDDHCRYMWTFLLEEESFAINEIFEEFSLVTQLSCAFLILRSDRGLEVSLENNL
                     CQESGIKRHLTAPYTPQQNGVVERRNRTLLEMCRSIMKHMHVPNYLWGEGIRHSTYLL
                     NRIATRSVQDKIPYELFMEKEPNISHLRIFGCICYAYVVSEVVTDLTEKWNNRSV
                     MEHSHNGNFILILGQPGNHGIGNEAETEETKHVENSEEDNDKDGQSFRKSKYLIDDYVLL
                     AEEEGERLLLCLNHEPFDFNEAEESEEWIQAFEEETHSIIENDTWNLMELPLCTEPIC
                     LKWVFKIKRNSDSSINKYMARLVAFCYVQFYGVDFQEVFAPVAELETFRCLLVNLVASN
                     GWETHHLDVKTAFLHGELSEEIYVVQPSGFEDPDHPDYVYKLNKALYGLKQASRAWNN
                     KLNRVLKELKFEKCSKEPSVYRKVMDGENILLVAVYVDDLFVIGTNLKLIDDFKRFMS
                     SKFFMSDLGPLTYYLGIEVCQHPDGIFISQDKYAKEILTRAGMEDCNPVHTPMEPNIK
                     LSKSAEENDIDATRFKKNPLYAKLKGVSQSCYEIMFKVCKRHIYIWENTWAFHEYDAK
                     PKLVGYSDRSHNVDPDDGRSTAGYIFYLGESSSLISWCSQKQDTVAMSSCEAEFMAGI
                     EAARQVIWLQDLLCEVTGLPCEKVVVRTDNKSAIALTKNLVFHGRNKHIHRKYHFIRE
                     CVENRHIEVEYVPGDIQFADILTKALGKVRFQVMRSLIGVQDLESSEEEVKGENVGLS
                     LKKT"
     gene            54086..55922
                     /gene "F28H19.7"
     CDS             join(54086..54446,54529..54822,55353..55922)
                     /gene "F28H19.7"
                     /codon_start 1
                     /product "Putative acyl-acyl carrier protein desaturase"
                     /protein_id "AAF63100.1"
                     /db_xref "GI:7523660"
                     /translation "MPSPSTPLASPFPGPAKISAVAAPVPPALKPQNKTHIMPPEKME
                     IFKSLDGWAKDQTLPLLKPVDQCWQPASFLPDPALPFSEPTDQVRELRERTASLPDEY
                     FVVLVGDMITEDALPTYQTMINTLDGVRDETGASESAWASWTRAWTAEENRHGDLLRT
                     YLYLSGRVDMLMVERTVQHLIGSGMDPCTENNPYLGFVYTSFQERATPVSHGNTARLA
                     KSAGDPVLARCTGTIAADEKKHENAYYVKVFKLLEIDPNGAVSAVADMMKKKIIMPAH
                     LMTDGRDPMLFEHFSAVAQRLEVYTADDYADILEFLVGRWRLEKLEGLTGEGQRAQEF
                     VCGLAQKTRRDQERADERAKKLKKTHEVCESWTFDKQISV"

alignment_scores:
              Quality:     7.00              Length:       7
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_66_74 x AC006423  ..

 Align seq 1/1  to: AC006423  from: 1  to: 123192

       1 IeuTyrProGluAspSerSer 7
         |||||||||||||||||||||
  109856 TTGTACCCGGAGGATTCATCC 109876

seq_name: qb btq25:CNS05TEV
<
```

```
seq documentation block:
LOCUS       CNS05TEV    134214 bp    DNA              HIG       08-DEC-2000
DEFINITION  Homo sapiens chromosome 14 clone c-222d18, *** SEQUENCING IN
            PROGRESS ***, in ordered pieces.
ACCESSION   AL559499
VERSION     AL559499.2  GI:11611186
KEYWORDS    HIG; HIGS PHASE2; HIGS DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 134214)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-DEC-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : Seqref@genoscope.cns.fr
            Web : www.genoscope.cns.fr)
COMMENT     On Dec 9, 2000 this sequence version replaced gi:8574317.
                  -- Genome Center
            Author: Genoscope / Centre National de Sequencage
            Center code: GS
            Web ...: http://www.genoscope.cns.fr/
            Contact: Seqref@genoscope.cns.fr
                 -- -- ---- -
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            plasmid, etc., even if efforts are made to eliminate these
            contaminating sequences. The following BAC sequence is oriented
            5'-3' the 17 -> the SP6 end.
            Beginning BAC (overlapping the SP6 end) : R-865E7
            Summary Statistics
            Assembly program: Phrap; version 2.0
            Consensus error: 35.71x in 920 bases, sum of contigs
                 -- ----- -- -
                    - -- --- ---------
            Overall quality chart :
            Range     :   bases
            0         :
            1     9   :       2
            10   19   :      24
            20   29   :     121
            30   39   :     513
            40   49   :    2011
            50   59   :    3307
            60   69   :    4771
            70   79   :   10305
            80   89   :   31755
            90   99   :   82405
                 --      -- ---
            Percentage of bases with a quality value > 40 : 99.9%.
            * NOTE: this is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..134214
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /clone="c-222d18"
                     /clone_lib="CalTech-D"
     STS             68473..68589
                     /note="matching EMBL:224616
                     RHdb:RH15514
                     RHdb:RH9724
                     dbSTS:SIS15402
                     Identified using the e-PCR software (G. Schuler)"
BASE COUNT     32284 a  41306 c  48152 g  24382 t
ORIGIN
```

```
alignment_scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_66_74 x CNS05TEV     .

 Align seq 1/1  to: CNS05TEV  from: 1  to: 134214

        8 ProGluAspSerSerTrpAla 9
          |||||||||||||||||||||
    67597 CCTGAGGACTCTTCCTGGGCC 67617

seq_name: qb_htq14:AC068677

seq_documentation_block:
LOCUS        AC068677    165613 bp    DNA              HIG       04-JUN-2000
DEFINITION   Homo sapiens chromosome 2 clone RP11 76114 map 2, WORKING DRAFT
             SEQUENCE, 14 unordered pieces.
ACCESSION    AC068677
VERSION      AC068677.2  GI:8247834
KEYWORDS     HIG; HIGS_PHASE1; HIGS_DRAFT.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 165613)
  AUTHORS    Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE      Homo sapiens chromosome 2, clone RP11-76114
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 165613)
  AUTHORS    Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
             Anderson,S., Baldwin,J., Barna,N., Bastien,V., Boda,F.,
             Boguslavkiy,I., Boukhgalter,B., Brown,A., Burkett,G.,
             Campopiano,A., Castle,A., Chevrot,Y., Colangelo,M., Collins,S.,
             Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
             Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
             Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
             Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
             Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
             Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
             Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
             McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
             Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mohan,V., Morrow,J.,
             Murphy,T., Naylor,J., Norman,C.B., O'Connor,T., O'Donnell,P.,
             O'Neil,D., Olivar,T.M., Oliver,J., Petersen,K., Pierre,N.,
             Pisani,C., Pollara,V., Raymond,C., Riley,R., Roper,P., Rothnagel,R.,
             Roy,A., Santos,R., Schauer,S., Searcy,P., Spencer,B.,
             Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
             Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Triquilio,J.,
             Vassiliev,H., Viel,R., Vo,A., Wilson,R., Wu,X., Wyman,D., Ye,W.J.,
             Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE      Direct Submission
  JOURNAL    Submitted (06-MAY-2000) Whitehead Institute/MIT Center for Genome
             Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT      On Jun 4, 2000 this sequence version replaced gi:7712142.
             All repeats were identified using RepeatMasker
             Smit, A.F.A. & Green, P. (1996-1997)
             http://ftp.genome.washington.edu:8080/RM/RepeatMasker.html
             -------------- Genome Center
                 Center: Whitehead Institute/MIT Center for Genome Research
                 Center code: WIBR
                 Web site: http://www.seq.wi.mit.edu
                 Contact: sequence_submissions@genome.wi.mit.edu
             ------- ------ Project Information
                 Center project name: LL0327
                 Center clone name: 76_I_14
             ------- ------ Summary Statistics
                 Sequencing vector: M13; M77815; 100% of reads
                 Chemistry: Dye terminator; Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.960731
                 Consensus quality: 159102 bases at least Q40
                 Consensus quality: 162057 bases at least Q30
                 Consensus quality: 163425 bases at least Q20
                 Insert size: 163000; agarose-tip
                 Insert size: 164313; sum-of-contigs
                 Quality coverage: 4.8 in Q20 bases; agarose-tip
                 Quality coverage: 4.7 in Q20 bases; sum-of-contigs
             ------------
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 14 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence
             * as soon as it is available and the accession number will
             * be preserved.
             *       1     1331: contig of 1331 bp in length
             *    1332  1431: gap of      100 bp
             *    1432     4176: contig of 2745 bp in length
             *    4177  4276: gap of      100 bp
             *    4277     7638: contig of 3362 bp in length
             *    7639  7738: gap of      100 bp
             *    7739    12691: contig of 4953 bp in length
             *   12692 12791: gap of      100 bp
             *   12792    18377: contig of 5586 bp in length
             *   18378 18477: gap of      100 bp
             *   18478    23070: contig of 4593 bp in length
             *   23071 23170: gap of      100 bp
             *   23171    30857: contig of 7687 bp in length
             *   30858 30957: gap of      100 bp
             *   30958    39987: contig of 9030 bp in length
             *   39988 40087: gap of      100 bp
             *   40088    49724: contig of 9637 bp in length
             *   49725 49824: gap of      100 bp
             *   49825    62946: contig of 13122 bp in length
             *   62947 63046: gap of      100 bp
             *   63047    83817: contig of 20771 bp in length
             *   83818 83917: gap of      100 bp
             *   83918   105840: contig of 21923 bp in length
             *  105841 105940: gap of      100 bp
             *  105941   125384: contig of 19444 bp in length
             *  125385 125484: gap of      100 bp
             *  125485   165613: contig of 40129 bp in length.
FEATURES             Location/Qualifiers
     source          1..165613
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11 76114"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..1331
                     /note="assembly_fragment"
     misc_feature    1432..4176
                     /note="assembly_fragment"
     misc_feature    4277..7638
                     /note="assembly_fragment"
     misc_feature    7739..12691
                     /note="assembly_fragment"
     misc_feature    12792..18377
                     /note="assembly_fragment"
     misc_feature    18478..23070
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    23171..30857
                     /note="assembly_fragment"
     misc_feature    30958..39987
                     /note="assembly_fragment"
     misc_feature    40088..49724
                     /note="assembly_fragment"
     misc_feature    49825..62946
                     /note="assembly_fragment"
```

```
     misc_feature    64047..88817
                     /note "assembly_fragment"
     misc_feature    88918..105840
                     /note "assembly_fragment"
     misc_feature    105941..125384
                     /note "assembly_fragment
                     clone end:SP6
                     vector_side:right"
     misc_feature    125485..165613
                     /note "assembly_fragment"
BASE COUNT    53167 a  28965 c  29519 g  52661 t    1301 others
ORIGIN


alignment scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_66_74 x AC068677    ..

 Align seq 1/1  to: AC068677  from: 1  to: 165613

       1 LeuTyrProGluAspSerSer 7
         ||||||||||||||||||||
  102166 CTTTACCCTGAAGACTCAAGT 102186

seq name: gb htg22:AL390787

seq documentation block:
LOCUS       AL390787    169329 bp    DNA              HTG       23-JAN-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-524L19, *** SEQUENCING IN
            PROGRESS ***, 24 unordered pieces.
ACCESSION   AL390787
VERSION     AL390787.4  GI:9926761
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169329)
  AUTHORS   Pavitt,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 25, 2000 this sequence version replaced gi:9717162.
              - - - - - - - Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
              - - - - - - Project Information
            Center project name: bA524L19
                  - - Summary Statistics
            Assembly program: X-GAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye terminator ET amersham; 45% of reads Chemistry:
            Dye terminator Big Dye; 64% of reads
            Consensus quality: 158972 bases at least Q40
            Consensus quality: 163198 bases at least Q30
            Consensus quality: 165163 bases at least Q20
            Insert size: 167129; sum of contigs
            Insert size: 188731; 5.6% error; approx fp
            Quality coverage: 3.55x in Q20 bases; sum of contigs quality
            coverage: 4.58x in Q20 bases; approx fp
                  .
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 24 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     3332: contig of  3332 bp in length
            *    3333  3432: gap of      100 bp
            *    3433     9796: contig of 6364 bp in length
            *    9797  9896: gap of      100 bp
            *    9897    18412: contig of 8516 bp in length
            *   18413 18512: gap of      100 bp
            *   18513    33107: contig of 14595 bp in length
            *   33108 33207: gap of      100 bp
            *   33208    42420: contig of 9213 bp in length
            *   42421 42520: gap of      100 bp
            *   42521    47055: contig of 4535 bp in length
            *   47056 47155: gap of      100 bp
            *   47156    52635: contig of 5480 bp in length
            *   52636 52735: gap of      100 bp
            *   52736    60939: contig of 8204 bp in length
            *   60940 61039: gap of      100 bp
            *   61040    65017: contig of 3978 bp in length
            *   65018 65117: gap of      100 bp
            *   65118    69318: contig of 4201 bp in length
            *   69319 69418: gap of      100 bp
            *   69419    73806: contig of 4388 bp in length
            *   73807 73906: gap of      100 bp
            *   73907    75995: contig of 2089 bp in length
            *   75996 76095: gap of      100 bp
            *   76096    81301: contig of 5206 bp in length
            *   81302 81401: gap of      100 bp
            *   81402    94129: contig of 12728 bp in length
            *   94130 94229: gap of      100 bp
            *   94230   101966: contig of 7737 bp in length
            *  101967 102066: gap of      100 bp
            *  102067   112736: contig of 10670 bp in length
            *  112737 112836: gap of      100 bp
            *  112837   125167: contig of 12331 bp in length
            *  125168 125267: gap of      100 bp
            *  125268   127720: contig of 2453 bp in length
            *  127721 127820: gap of      100 bp
            *  127821   137241: contig of 9421 bp in length
            *  137242 137341: gap of      100 bp
            *  137342   144614: contig of 7273 bp in length
            *  144615 144714: gap of      100 bp
            *  144715   149623: contig of 4909 bp in length
            *  149624 149723: gap of      100 bp
            *  149724   162430: contig of 12707 bp in length
            *  162431 162530: gap of      100 bp
            *  162531   169329: contig of 6799 bp in length.
FEATURES             Location/Qualifiers
     source          1..169329
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /chromosome "1"
                     /clone "RP11-524L19"
                     /clone_lib "RPCI-11.2"
     misc_feature    1..3332
                     /note "assembly_fragment:00205
                     clone end:SP6
                     vector_side:left"
     misc_feature    3433..9796
                     /note "assembly_fragment:00029"
     misc_feature    9897..18412
                     /note "assembly_fragment:00126"
     misc_feature    18513..33107
                     /note "assembly_fragment:00129"
     misc_feature    33208..42420
                     /note "assembly_fragment:00136"
     misc_feature    42521..47055
                     /note "assembly_fragment:00156.0"
     misc_feature    47156..52635
                     /note "assembly_fragment:00213"
     misc_feature    52736..60939
                     /note "assembly_fragment:00283"
```

```
     misc_feature    61940. .65017
                     /note="assembly_fragment:00600"
     misc_feature    65118. .69418
                     /note="assembly_fragment:00692"
     misc_feature    69419. .73806
                     /note="assembly_fragment:00716"
     misc_feature    73907. .75965
                     /note="assembly_fragment:00784"
     misc_feature    76096. .81401
                     /note="assembly_fragment:00821"
     misc_feature    81402. .94129
                     /note="assembly_fragment:00887"
     misc_feature    94230. .101966
                     /note="assembly_fragment:00994"
     misc_feature    102067. .112736
                     /note="assembly_fragment:01104"
     misc_feature    112837. .125167
                     /note="assembly_fragment:01130"
     misc_feature    125268. .127720
                     /note="assembly_fragment:01141"
     misc_feature    127821. .137241
                     /note="assembly_fragment:01225"
     misc_feature    137342. .144614
                     /note="assembly_fragment:01255"
     misc_feature    144715. .149623
                     /note="assembly_fragment:01310"
     misc_feature    149724. .162430
                     /note="assembly_fragment:01358"
     misc_feature    162531. .169329
                     /note="assembly_fragment:01491"
BASE COUNT    49992 a   33544 c   33720 g   49860 t    2213 others
ORIGIN


alignment_scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_66_74 x AL390787   ..

 Align seq 1/1  to: AL390787  from: 1  to: 166329

       1 LeuTyrProGluAspSerSer 7
         |||||||||||||||||||||
  147531 CTATACCCAGAAGATAGCAGC 147551

seq_name: qb_htgl:AC007608

seq_documentation_block:
LOCUS       AC007608   170757 bp    DNA             HTG       13-JUL-2000
DEFINITION  Homo sapiens chromosome 16 clone RP11-401P9, WORKING DRAFT
            SEQUENCE, 8 ordered pieces.
ACCESSION   AC007608
VERSION     AC007608.3  GI:8575872
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 170757)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 170757)
  AUTHORS   Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,
            Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,
            Bryant,J., Tesmer,J., Meincke,L., Longmire,J., White,S., Tatum,O.,
            Campbell,C., Fawcett,J., Maltbie,M., Bussod,M., Sutherland,R.,
            McMurry,K., Han,C. and Deaven,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAY-1999) Center for Human Genome Studies, DOE Joint
            Genome Institute, Los Alamos National Laboratory, MS M888, Los
            Alamos, NM 87545, USA
COMMENT     On Jun 21, 2000 this sequence version replaced gi:6624965.
              ---Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
              ---
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.
              ---------- ------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submitter.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *         1    18718: contig of 18718 bp in length
            *     18719    18818: gap of unknown length
            *     18819    41015: contig of 22197 bp in length
            *     41016    41115: gap of unknown length
            *     41116    43507: contig of 2392 bp in length
            *     43508    43607: gap of unknown length
            *     43608    45562: contig of 1955 bp in length
            *     45563    45662: gap of unknown length
            *     45663   121631: contig of 75969 bp in length
            *    121632   121731: gap of unknown length
            *    121732   151445: contig of 29714 bp in length
            *    151446   151545: gap of unknown length
            *    151546   164045: contig of 12500 bp in length
            *    164046   164145: gap of unknown length
            *    164146   170757: contig of 6612 bp in length.
FEATURES             Location/Qualifiers
     source          1. .170757
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /chromosome="16"
                     /clone "RP11-401P9"
BASE COUNT    44099 a   45003 c   41198 g   39751 t     706 others
ORIGIN


alignment_scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_66_74 x AC007608   ..

 Align seq 1/1  to: AC007608  from: 1  to: 170757

       1 LeuTyrProGluAspSerSer 7
         |||||||||||||||||||||
  114572 CTCTACCCGGAAGACTCTTCC 114592
```

OM of: US-09-126-945B-2_COPY_66_74 [redacted]

Date: Sep 14, 2001, 9:48 PM

About: Results were produced by the [redacted]
Copyright (c) [redacted]

Command line parameters:
-MODEL=frame+_p2n.model -DEV_xlp
-Q [redacted]
-DB=N_Geneseq_0601 -QFMT=fastap -SUFFIX=olip2n.rng -GAPOP 4.500
-GAPEXT 0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT 0.000
-QGAPOP 4.500 -QGAPEXT=0.050 -XGAPOP 60.000 -XGAPEXT=60.000
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
-DELOP=6.000 -DELEXT 7.000 -START=1 -MATRIX otiqo
-TRANS human40.otl -LIST=45 -DOCALIGN=200 -THR_SCORE=quality
-THR_MIN 1 -ALIGN 15 -MODE LOCAL -OUTFMT pir -NORM ext -MINGEN 9
-MAXLEN 2000000000 -USER=US09126945_*COGNI_1_08 -NCPU 6 -TCPU 4
-LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-126-945B-2_COPY_66_74
Query length: 9
Database: N_Geneseq_0601:*
Database sequences: 740101
Database length: 313950809
Search time (sec): 680.150000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation	..			
/SIDSl/seqdata/geneseq/geneseqn/NA2000.DAT:AAF21828	-	9	9.00	171.66	0.1366	1087	!			
/SIDSl/seqdata/geneseq/geneseqn/NA2000.DAT:AAF21614	-	9	9.00	171.66	0.1366	1961	!			
/SIDSl/seqdata/geneseq/geneseqn/NA2001.DAT:AAC83261	-	9	9.00	171.66	0.1366	1894	!			
/SIDSl/seqdata/geneseq/geneseqn/NA2001.DAT:AAC83266	-	9	9.00	167.73	0.2261	3417	!			
/SIDSl/seqdata/geneseq/geneseqn/NA1994.DAT:AAQ67954	-	6	6.00	143.40	5.13	21	!			
/SIDSl/seqdata/geneseq/geneseqn/NA1994.DAT:AAQ55004	-	6	6.00	143.40	5.13	21	!			
/SIDSl/seqdata/geneseq/geneseqn/NA2000.DAT:AAC30610	-	6	6.00	130.73	26.04	128	!			
/SIDSl/seqdata/geneseq/geneseqn/NA1995.DAT:AAQ91962	-	6	6.00	130.51	26.39	132	!			
/SIDSl/seqdata/geneseq/geneseqn/NA2000.DAT:AAC74665	-	6	6.00	126.49	12.56	221	!			
/SIDSl/seqdata/geneseq/geneseqn/NA2000.DAT:AAC74675	-	6	6.00	122.19	75.01	415	!			
/SIDSl/seqdata/geneseq/geneseqn/NA2000.DAT:AAC16761	-	6	6.00	117.65	139.30	876	!			
/SIDSl/seqdata/geneseq/geneseqn/NA1999.DAT:AAX75523	-	6	6.00	117.30	145.65	968	!			
/SIDSl/seqdata/geneseq/geneseqn/NA1997.DAT:AAT77777	-	6	6.00	117.05	150.48	998	!			
/SIDSl/seqdata/geneseq/geneseqn/NA1997.DAT:AAT66004	-	6	6.00	117.05	150.48	998	!			
/SIDSl/seqdata/geneseq/geneseqn/NA2000.DAT:AAA72161	-	6	6.00	116.51	160.00	998	!			
/SIDSl/seqdata/geneseq/geneseqn/NA1992.DAT:AAZ14368	-	6	6.00	116.21	167.51	1011	!			
/SIDSl/seqdata/geneseq/geneseqn/NA1997.DAT:AAT77772	-	6	6.00	115.97	172.85	1050	!			
/SIDSl/seqdata/geneseq/geneseqn/NA1997.DAT:AAT77773	-	6	6.00	115.97	172.85	1050	!			
/SIDSl/seqdata/geneseq/geneseqn/NA1997.DAT:AAT60094	-	6	6.00	115.97	172.85	1050	!			
/SIDSl/seqdata/geneseq/geneseqn/NA1997.DAT:AAT60090	-	6	6.00	115.97	172.85	1050	!			
/SIDSl/seqdata/geneseq/geneseqn/NA2000.DAT:AAC45208	-	6	6.00	114.88	198.70	1226	!			
/SIDSl/seqdata/geneseq/geneseqn/NA2000.DAT:AAC13710	-	6	6.00	114.54	207.71	1288	!			
/SIDSl/seqdata/geneseq/geneseqn/NA1997.DAT:AAT79902	-	6	6.00	114.47	238.22	1500	!			
/SIDSl/seqdata/geneseq/geneseqn/NA1998.DAT:AAV76873	-	6	6.00	114.47	238.22	1500	!			
/SIDSl/seqdata/geneseq/geneseqn/NA2000.DAT:AAY59314	-	6	6.00	113.22	245.72	1563	!			
/SIDSl/seqdata/geneseq/geneseqn/NA2000.DAT:AAC75849	-	6	6.00	112.76	260.95	1660	!			
/SIDSl/seqdata/geneseq/geneseqn/NA2001.DAT:AAA61011	-	6	6.00	111.11	332.40	2100	!			
/SIDSl/seqdata/geneseq/geneseqn/NA1999.DAT:AAA30650	-	6	6.00	110.58	344.95	2204	!			
/SIDSl/seqdata/geneseq/geneseqn/NA1998.DAT:AAV14251	-	6	6.00	110.21	361.76	2307	!			
/SIDSl/seqdata/geneseq/geneseqn/NA1998.DAT:AAV19253	-	6	6.00	109.35	454.02	2911	!			
/SIDSl/seqdata/geneseq/geneseqn/NA1999.DAT:AAY87705	-	6	6.00	109.35	454.02	2911	!			
/SIDSl/seqdata/geneseq/geneseqn/NA2000.DAT:AAC00701	-	6	6.00	109.35	454.02	2911	!			
/SIDSl/seqdata/geneseq/geneseqn/NA1999.DAT:AAZ01425	-	6	6.00	67.61	8.6e-001	1980	!			
/SIDSl/seqdata/geneseq/geneseqn/NA1991.DAT:AAZ14430	-	5	5.00	123.16	66.26	21	!			
/SIDSl/seqdata/geneseq/geneseqn/NA1993.DAT:AAQ24902	-	5	5.00	122.23	77.40	25	!			
/SIDSl/seqdata/geneseq/geneseqn/NA1999.DAT:AAX89202	-	5	5.00	121.69	82.94	27	!			
/SIDSl/seqdata/geneseq/geneseqn/NA2001.DAT:AAZ01218	-	5	5.00	121.69	82.94	27	!			
/SIDSl/seqdata/geneseq/geneseqn/NA1997.DAT:AAT70256	-	5	5.00	120.24	99.45	33	!			
/SIDSl/seqdata/geneseq/geneseqn/NA2000.DAT:AAC97823	-	5	5.00	119.72	105.24	42	!			
/SIDSl/seqdata/geneseq/geneseqn/NA1997.DAT:AAV76964	-	5	5.00	115.32	187.82	67	!			
/SIDSl/seqdata/geneseq/geneseqn/NA1999.DAT:AAX51878	-	5	5.00	112.95	254.68	94	!			
/SIDSl/seqdata/geneseq/geneseqn/NA1998.DAT:AAC94857	-	5	5.00	112.31	276.50	104	!			
/SIDSl/seqdata/geneseq/geneseqn/NA1997.DAT:AAN71261	-	5	5.00	111.91	290.95	109	!			
/SIDSl/seqdata/geneseq/geneseqn/NA2000.DAT:AAF13097	-	5	5.00	111.72	298.14	112	!			
/SIDSl/seqdata/geneseq/geneseqn/NA1995.DAT:AAL23239	-	5	5.00	111.54	305.32	114	!			

Alignment for: /SIDSl/seqdata/geneseq/geneseqn/NA2000.DAT:AAF21828

Seq_documentation_block:
ID AAF21828 standard; DNA; 1087 BP.
XX
AC AAF21828;
XX
DT 27-MAR-2001 (first entry)
XX
DE Human breast and ovarian cancer associated antigen gene SEQ ID 215.
XX
KW Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
KW nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
KW antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
KW Addison's disease; allergy; autoimmune haemolytic anaemia;
KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW cardiovascular disorder; wound healing; neurological disease; ds.
XX
OS Homo sapiens.
XX
PN WO200055173-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05881.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA) HUMAN GENOME SCI INC.
XX
PI Rosen CA; Ruben SM;
XX
DR WPI; 2000-611515/58.
DR P-PSDB; AAB58925.
XX
PT New human breast and ovarian cancer associated gene sequences and the
PT polypeptides encoded by these genes, useful in the prevention,
PT treatment and diagnosis of cancer, immune disorders, cardiovascular
PT disorders and neurological diseases -
XX
PS Claim 1; Page 442-613; 1299pp; English.
XX
CC Sequences AAF21614 - AAF22041 represent DNA sequences encoding human
CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
CC associated with breast and ovarian cancer. Included in the invention are
CC sequences AAF21042 - AAF22040 and AAB59129 which are used in the
CC isolation and characterisation of the DNA and protein sequences of the
CC invention. The breast and ovarian cancer associated DNA, protein, agonist
CC or antagonist sequences exhibit cytostatic; immunosuppressive;
CC nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
CC antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
CC antibacterial; antifungal; antiparasitic and cardiant activity. The
CC polynucleotide and protein sequences are used in the diagnosis, cancer,
CC particular, breast and ovarian cancer. The nucleic acid sequences,
CC probes, arrays and chips may also be used in the diagnosis,
CC prevention and treatment of immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis, cardiovascular disorders such as
CC myocardial ischaemia, wound healing, neurological diseases such as
CC cerebral ischaemia, epilepsy, and infectious diseases
XX
SQ Sequence 1087 BP; 215 A; 347 C; 349 G; 192 T; 4 other;

```
alignment scores:
            Quality:     9.00              Length:       9
              Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2-COPY_66_74 x AAZ21828    ..

 Align seq 1/1   to: AAZ21828   from: 1   to: 1067

        1 LeuTyrProGluAspSerSerTrpAla 9
          |||||||||||||||||||||||||||
       72 CTGTACCCTGACGACAGCAGCTGGGCA 98

Seq_name: /cgn2_6/prod/1pa_gb/seq_gb/gbpri3.seq:gbpri3:AAZ50691

seq documentation block:
ID   AAZ50691 standard; cDNA; 1894 BP.
XX
AC   AAZ50691;
XX
DT   23-MAY-2000  (first entry)
XX
DE   Human prostate derived Ets factor cDNA.
XX
KW   Prostate Derived Ets factor; PDEF; human; chromosome 6p21.3; cancer;
KW   loss of heterozygosity; chromosomal translocation; linkage analysis;
KW   cytostatic; cardiant; immunosuppressive; cerebroprotective; fungicide;
KW   antibacterial; vulnerary; neuroprotective; antiparkinsonian; nootropic;
KW   anabolic; antiinflammatory; anorectic; hybridisation probe; forensic;
KW   tumour marker, diagnosis, treatment, prostate cancer, blood coagulation,
KW   autoimmune disorder; haematopoietic; immune/nervous system; stroke;
KW   neoplasm; microbial infection; tissue regeneration; heart attack;
KW   scarring; food additive; preservative; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             416..1423
FT                   /*tag= a
FT                   /product= "Human prostate derived Ets factor"
FT                   /note= "Expressed in prostate epithelium"
XX
PN   WO200006589-A1.
XX
PD   10-FEB-2000.
XX
PF   02-AUG-1999;   99WO-US17470.
XX
PR   31-JUL-1998;   98US-0126945.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI   Libermann TA, Oettgen JP, Kunsch CA, Endress GA, Rosen CA;
XX
DR   WPI; 2000-195255/17.
DR   P-PSDB; AAY45003.
XX
PT   Novel prostate derived polypeptide, polynucleotide useful for
PT   diagnosis, prevention and treatment of prostate cancer, autoimmune
PT   disorders, microbial infections and also as food additive or
PT   preservative
XX
PS   Claim 4; Fig 1; 132pp; English.
XX
CC   The present cDNA sequence encodes a human prostate derived Ets factor
CC   (PDEF), isolated from human prostate carcinoma cDNA library
CC   (ATCC No.203072). PDEF gene expression is highest in tissues involved in
CC   androgen sensitivity, with enriched expression in prostate epithelium
CC   cells. PDEF gene is mapped to the human chromosome 6p21.3
CC   region that is associated with loss of heterozygosity and chromosomal
CC   translocations in various human cancers. PDEF has cytostatic, cardiant,
CC   immunosuppressive, cerebroprotective, fungicide, antibacterial,
CC   vulnerary, neuroprotective, antiparkinsonian, nootropic, anabolic,
CC   antiinflammatory and anorectic activity. PDEF polynucleotides are useful
CC   in linkage analysis as markers, as hybridisation probes for differential
CC   identification of the tissues or cell types and as polymorphic markers
CC   for forensic purposes. PDEF is useful as prostate specific tumour marker
CC   for the diagnosis and treatment of prostate cancer. PDEF sequences are
CC   useful for treating autoimmune disorders, haematopoietic, blood
CC   coagulation, immune and nervous system disorders, hyperproliferative
CC   disorders like, neoplasms and microbial infections, heart attacks,
CC   stroke, scarring and for tissue regeneration. They are also useful as
CC   food additives or preservatives.
XX
SQ   Sequence 1894 BP; 368 A; 653 C; 571 G; 302 T; 0 other;


alignment scores:
            Quality:     9.00              Length:       9
              Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2-COPY_66_74 x AAZ50691    ..

 Align seq 1/1   to: AAZ50691   from: 1   to: 1894

        1 LeuTyrProGluAspSerSerTrpAla 9
          |||||||||||||||||||||||||||
      611 CTGTACCCTGACGACAGCAGCTGGGCA 637

Seq_name: /cgn2_6/prod/1pa_gb/seq_gb/gbpri3.seq:gbpri3:AAC83261

seq documentation block:
ID   AAC83261 standard; DNA; 1894 BP.
XX
AC   AAC83261;
XX
DT   16-MAR-2001  (first entry)
XX
DE   Gland specific Ets transcription factor (GSEF) nucleotide sequence.
XX
KW   Transcription factor; gland-specific Ets transcription factor; GSEF;
KW   metastatic potential; cancer; tumour; metastasis; breast; prostate;
KW   leukaemia; lymphoma; sarcoma; melanoma; chromosome 6p21.1-6p21.3; ds.
XX
OS   Homo sapiens.
XX
PN   WO200070092-A1.
XX
PD   23-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US13174.
XX
PR   14-MAY-1999;   99US-0134112.
XX
PA   (CHIR-) CHIRON CORP.
XX
PI   Kaufmann J, Xin H, Barrowe G;
XX
DR   WPI; 2001-041015/05.
DR   P-PSDB; AAB49628.
XX
PT   Detecting metastatic and potential metastatic cancerous cells, useful
PT   for diagnosing, prognosing, grading and staging of cancers by detecting
PT   gland specific Ets transcription factor gene product in a biological
PT   sample from a cell
XX
PS   Disclosure; Page 81-83; 95pp; English.
XX
CC   This invention relates to a method for the detection and determination of
CC   the metastatic potential of a cell. The method comprises detecting a
CC   gland-specific Ets transcription factor (GSEF) gene product in a test
```

```
CC   sample. Detection of a GSEF gene product in the test sample in amount
CC   lower than that in a normal cell, is indicative of a cell with high
CC   metastatic potential. The method is useful for determining the metastatic
CC   potential of a cell, for the diagnosis and prognosis of cancer as well as
CC   grading and staging of cancers by detecting GSEF expression in a
CC   biological test sample. The method may also be used to monitor patients
CC   having a predisposition to develop a particular cancer. GSEF polypeptides
CC   are useful for producing antibodies, in cancer diagnosis, prognosis,
CC   grading, staging and management of breast and prostate tumours, and in
CC   detecting polymorphisms in the sequence. GSEF genes and proteins are also
CC   useful in gene therapy. GSEF gene product expression levels can be used
CC   in conjunction with any tissue in which an alteration in GSEF gene
CC   product expression levels is associated with development of a
CC   cancer associated phenotype. Cancers, which can be monitored include
CC   cancers of the prostate, cervix, lung and colon, melanomas, colorectal
CC   adenocarcinomas, Wilm's tumour, retinoblastoma, sarcomas, myosarcomas,
CC   lung carcinomas, leukaemia, and lymphomas. The GSEF gene is located on
CC   human chromosome 6, specifically at 6p21.1 6p21.3. The present sequence
CC   represents the DNA encoding GSEF
XX
SQ   Sequence 1894 BP; 368 A; 653 C; 571 G; 302 T; 0 other;


alignment_scores:
            Quality:    9.00            Length:      9
              Ratio:   1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_66_74 x AAC83261    ..

 Align seq 1/1  to: AAC83261  from: 1  to: 1894

       1 LeuTyrProGluAspSerSerTrpAla 9
         ||||||||||||||||||||||||||
     611 CTGTACCCTGAGGACAGCAGCTGGGCA 637

seq_name: /SIDS1/5 ;data/geneseq/geneseqp/NA2*.dat AAW7*

seq_documentation_block:
ID   AAC83266 standard; DNA; 1817 BP.
XX
AC   AAC83266;
XX
DT   16-MAR-2001  (first entry)
XX
DE   Gland-specific Ets transcription factor (GSEF) cDNA sequence.
XX
KW   Transcription factor; gland-specific Ets transcription factor; GSEF;
KW   metastatic potential; cancer; tumour; metastasis; breast; prostate;
KW   leukaemia; lymphoma; sarcoma; melanoma; chromosome 6p21.1 6p21.3; ss.
XX
OS   Homo sapiens.
XX
PN   WO200070092-A1.
XX
PD   23-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US13173.
XX
PR   14-MAY-1999;   99US-0134112.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Kaufmann J,  Xin H,  Harrowe G;
XX
DR   WPI; 2001-041019/05.
DR   P-PSDB; AAB49633.
XX
PT   Detecting metastatic and potential metastatic cancerous cells, useful
PT   for diagnosing, prognosing, grading and staging of cancers by detecting
PT   gland-specific Ets transcription factor gene product in a biological
PT   sample from a cell -
XX
PS   Disclosure; Fig 2; 95pp; English.
XX
CC   This invention relates to a method for the detection and determination of
CC   the metastatic potential of a cell. The method comprises detecting a
CC   gland-specific Ets transcription factor (GSEF) gene product in a test
CC   sample. Detection of a GSEF gene product in the test sample in amount
CC   lower than that in a normal cell, is indicative of a cell with high
CC   metastatic potential. The method is useful for determining the metastatic
CC   potential of a cell, for the diagnosis and prognosis of cancer as well as
CC   grading and staging of cancers by detecting GSEF expression in a
CC   biological test sample. The method may also be used to monitor patients
CC   having a predisposition to develop a particular cancer. GSEF polypeptides
CC   are useful for producing antibodies, in cancer diagnosis, prognosis,
CC   grading, staging and management of breast and prostate tumours, and in
CC   detecting polymorphisms in the sequence. GSEF genes and proteins are also
CC   useful in gene therapy. GSEF gene product expression levels can be used
CC   in conjunction with any tissue in which an alteration in GSEF gene
CC   product expression levels is associated with development of a
CC   cancer-associated phenotype. Cancers, which can be monitored include
CC   cancers of the prostate, cervix, lung and colon, melanomas, colorectal
CC   adenocarcinomas, Wilm's tumour, retinoblastoma, sarcomas, myosarcomas,
CC   lung carcinomas, leukaemia, and lymphomas. The GSEF gene is located on
CC   human chromosome 6, specifically at 6p21.1 6p21.3. The present sequence
CC   represents the cDNA encoding GSEF.
XX
SQ   Sequence 1817 BP; 710 A; 1026 C; 970 G; 611 T; 0 other;


alignment_scores:
            Quality:    9.00            Length:      9
              Ratio:   1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_66_74 x AAC83266    ..

 Align seq 1/1  to: AAC83266  from: 1  to: 1817

       1 LeuTyrProGluAspSerSerTrpAla 9
         ||||||||||||||||||||||||||
    2010 CTGTACCCTGAGGACAGCAGCTGGGCA 2036

seq_name: /SIDS1/5 ;data/geneseq/geneseqp/NA1*.dat AAQ*

seq_documentation_block:
ID   AAQ67954 standard; DNA; 21 BP.
XX
AC   AAQ67954;
XX
DT   08-DEC 1994  (first entry)
XX
DE   CD44 PCR primer ES.
XX
KW   CD44; exon 6; cell surface protein; primer; polymerase chain reaction;
KW   PCR; amplification; tumor; imaging; therapy; ss.
XX
OS   Synthetic.
XX
PN   WO9412631-A.
XX
PD   09-JUN 1994.
XX
PF   22-NOV 1993;   93WO-GB02394.
XX
PR   20-NOV 1992;   92GB-0024386.
PR   16-DEC 1992;   92GB-0026165.
PR   20-JUL 1993;   93WO-GB01520.
XX
PA   (ISIS-) ISIS INNOVATION LTD.
XX
PI   Matsumura Y.  Iarin D;
XX
```

DR WPI; 1994-200260/24.
XX
PT New antibody specific for peptide of exon 4 of CD44 - also exon 6
PT peptide fragments and fusion protein, useful for diagnosis,
PT treatment and imaging of tumours
XX
PS Disclosure; Page 25; 74pp; English.
XX
CC The primers given in AAQ67948-58 are used for cancer diagnosis by
CC PCR assay. The primers given in AAQ67948-51 anneal to the standard
CC protein of CD44, and those given in AAQ67952-58 anneal to exon
CC sequences.
XX
SQ Sequence 21 BP; 5 A; 6 C; 7 G; 3 T; 0 other;

alignment scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2-COPY-66-74 x AAQ67954 .

Align seq 1/1 to: AAQ67954 from: 1 to: 21

3 ProGluAspSerSerTrp 8
4 CCAGAGGACAGTTCCTGG 21

Seq name: /cgn2_6/prodata/geneseq/geneseqp-pc/NA2000.DAT:AAQ67954 1

Seq documentation block
ID AAQ55904 standard; DNA; 21 BP.
XX
AC AAQ55904;
XX
DT 25-JUL-1994 (first entry)
XX
DE PCR primer for amplifying region of CD44 gene.
XX
KW PCR; polymerase chain reaction; CD44; cancer; bladder cancer;
KW detection; diagnosis; carcinoma; solid tumour; ss.
XX
OS Synthetic.
XX
PN WO9402644-A.
XX
PD 03-FEB-1994.
XX
PF 20-JUL-1993; 93WO-GB01520.
XX
PR 21-JUL-1992; 92GB-0015498.
PR 20-NOV-1992; 92GB-0024386.
PR 16-DEC-1992; 92GB-0026165.
XX
PA (ISIS) ISIS INNOVATION LTD.
XX
PI Barin D, Matsumura Y;
XX
DR WPI; 1994-048890/06.
XX
PT Method for diagnosing neoplasia, e.g. carcinoma - utilises
PT unequal splicing of CD44 gene in cancer patients compared to
PT controls
XX
PS Example 4; Page 17; 42pp; English.
XX
CC 11 primers (AAQ55898-Q55908) were used in the amplification of a CD44
CC gene region and in the subsequent detection of the amplification
CC products. cDNA derived from urine of 90 patients (44 with biopsy
CC proven bladder cancer, 46 from normal volunteers and patients with
CC neoplastic inflammation of the bladder) was divided equally into two
CC tubes, one for PCR with primers E1 (AAQ55902) and E5 (AAQ55906) to
CC amplify the transcript for diagnosis, the other for PCR with P1
CC (AAQ55899) and P4 (AAQ55901) to amplify the standard form of CD44.
CC When all exons were expressed (bladder cancer), E1 and E4 produced a
CC 735 bp band. There was no band in tracks containing cDNA from
CC normal patients or those with bladder inflammation. A 482 bp band
CC obtained in all cases with P1 and P4 indicated that diagnostically
CC significant differences between urine from patients with bladder
CC cancer and controls was due to alternative splicing of the CD44
CC gene. The method allows diagnosis of neoplasia based on
CC over expression of the CD44 exon. Solid tumours, especially
CC carcinomas, can be accurately detected. This primer is
CC complementary to a sequence contained within exon 11 of the CD44
CC gene.
XX
SQ Sequence 21 BP; 5 A; 6 C; 7 G; 3 T; 0 other;

alignment scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2_COPY-66-74 x AAQ55904 ..

Align seq 1/1 to: AAQ55904 from: 1 to: 21

3 ProGluAspSerSerTrp 8
4 CCAGAGGACAGTTCCTGG 21

Seq name: /cgn2_6/prodata/geneseq/geneseqp-pc/NA2000.DAT:AAC30610

Seq documentation block:
ID AAC30610 standard; cDNA; 128 BP.
XX
AC AAC30610;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 84685.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1043401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 9905-0122487.
XX
PA (GESF) GENSEL.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500481/45.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures
XX
PS Claim 1; SEQ ID 84685; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)

CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
SQ Sequence 128 BP; 22 A; 36 C; 29 G; 40 T; 1 other;

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US 09-126-945B-2_COPY_66_74 x AAQ81962 ..

Align seq 1/1 to reverse of: AAQ81962 from: 1 to: 128

3 ProGluAspSerSerTrp 8
|||||||||||||||||||
52 CCAGAGGACTCCTCCTGG 35

seq_name: /cgi/e1/geneseq/geneseqp_NA2.ofr:EAT:AAQ81962

seq_documentation_block:
ID AAQ81962 standard; DNA; 132 BP.
XX
AC AAQ81962;
XX
DT 08-SEP-1995 (first entry)
XX
DE Human CD44 variable exon v7.
XX
KW Tumour; CD44; epitope; variable exon; gastric; colonic; mammary;
KW cervical; carcinoma; detection; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT exon 1..132
FT /*tag= a
FT /number= v7
FT /note= "reading frame continues from position 3"
XX
PN WO9500851-A.
XX
PD 05-JAN-1995.
XX
PF 15-JUN-1994; 94WO-EP01952.
XX
PR 22-JUN-1993; 93DE-4320624.
PR 22-JUN-1993; 93DE-4320624.
PR 02-JUL-1993; 93DE-4321944.
PR 28-APR-1994; 94DE-4414787.
XX
PA (BOEH) BOEHRINGER INGELHEIM INT GMBH.
XX
PI Heider K, Herrlich P, Ponta H, Dall P, Pals ST;
XX
DR WPI; 1995-052230/07.
DR P-PSDB; AAR69550.
XX
PT Analysis and detection of tumours using variable CD44 exons - in
PT particular CD44 exons v5, v6, v7, v8 or a combination of v7 and
PT v8; useful in the analysis of mammary, colonic and cervical
PT carcinoma.
XX
PS Claim 5; Page 40; 59pp; German.

XX
CC Tumours can be detected by detecting at least one variable exon of
CC the CD44 gene or a corresponding epitope. The variable epitopes are
CC esp. v5, v6, v7 or v8 (see AAQ81960-AAQ81964). These exons (and antibodies
CC against the epitopes they encode) are useful in the diagnosis and
CC analysis of tumours, esp. mammary, stomach, colonic and cervical
CC carcinomas.
XX
SQ Sequence 132 BP; 42 A; 40 C; 30 G; 20 T; 0 other;

alignment scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US 09-126-945B-2_COPY_66_74 x AAQ81962 ..

Align seq 1/1 to: AAQ81962 from: 1 to: 132

3 ProGluAspSerSerTrp 8
|||||||||||||||||||
51 CCAGAGGACTCCTCCTGG 68

seq_name: /cgi/e1/geneseq/geneseqp_NA2.ofr:EAT:AAC02665

seq_documentation_block:
ID AAC02665 standard; cDNA; 221 BP.
XX
AC AAC02665;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 2663.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99DE-0122487.
XX
PA (GESI) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
DR P-PSDB; AAG02659.
XX
PT Novel nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'-ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures
XX
PS Claim 1; SEQ ID 2663; 71pp + CD ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or poly A+ RNAs
CC derived from 40 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,

CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
SQ Sequence 221 BP; 54 A; 42 C; 69 G; 54 T; 3 other;

alignment scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2_COPY_66_74 x AAC74675/rev

Align seq 1/2 to reverse of: AAC74675 from: 1 to: 221

2 TyrProGluAspSerSer 7
|||||||||||||||||||
180 TACCCAGAGGATTCCTCT 163

Seq name: /tic2/1:data_gene_seq_search_pc_NA2...:AAC74675

Seq documentation block:
ID AAC74675 standard; cDNA; 415 BP.
XX
AC AAC74675;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF230 polynucleotide sequence SEQ ID NO:459.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540764.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Lough M;
XX
DR WPI; 2000-609262/57.
DR P-PSDB; AAB40466.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease
XX
PS Claim 5; Page 660; 5507pp; English.
XX

CC AAC74414 to AAC77606 encode the proteins given in AAB40237 to AAB43597,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 415 BP; 81 A; 128 C; 120 G; 86 T; 0 other;

alignment scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2_COPY_66_74 x AAC74675/rev

Align seq 1/2 to reverse of: AAC74675 from: 1 to: 415

4 GluAspSerSerTrpAla 9
|||||||||||||||||||
295 GAAGATTCGTCCTGGGCT 278

Seq_name: /tic2/1:data_gene_seq_search_pc_NA2000.DAT:AAC46754

Seq_documentation_block:
ID AAC46754 standard; DNA; 826 BP.
XX
AC AAC46754;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 51280.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0124548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 24-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142804.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145914.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147304.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152364.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157758.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.

```
PR   14 OCT 1999;   99US-0159341.
PR   14 OCT 1999;   99US-0159647.
PR   14 OCT 1999;   99US-0159648.
PR   18 OCT 1999;   99US-0159584.
PR   21 OCT 1999;   99US-0160741.
PR   21 OCT 1999;   99US-0160767.
PR   21 OCT 1999;   99US-0160768.
PR   21 OCT 1999;   99US-0160770.
PR   21 OCT 1999;   99US-0160814.
PR   21 OCT 1999;   99US-0160815.
PR   22 OCT 1999;   99US-0160980.
PR   22 OCT 1999;   99US-0160981.
PR   22 OCT 1999;   99US-0160989.
PR   25 OCT 1999;   99US-0161404.
PR   25 OCT 1999;   99US-0161405.
PR   25 OCT 1999;   99US-0161406.
PR   26 OCT 1999;   99US-0161359.
PR   26 OCT 1999;   99US-0161360.
PR   26 OCT 1999;   99US-0161361.
PR   28 OCT 1999;   99US-0161920.
PR   28 OCT 1999;   99US-0161992.
PR   28 OCT 1999;   99US-0161993.
PR   29 OCT 1999;   99US-0162142.

alignment scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2 COPY_66_74 x AAC46754   ..

 Align seq 1/1  to: AAC46754  from: 1  to: 826

       2 TyrProGluAspSerSer 7
         ||||||||||||||||||
     216 TACCCGGAAGACTCGTCC 233

seq names: /cgbin/..prdata/geneseq.geneseqp_NA1999.DAT,AAX37523

seq documentation block:
ID   AAX37523 standard; cDNA; 868 BP.
XX
AC   AAX37523;
XX
DT   06-JUL-1999  (first entry)
XX
DE   Human secreted protein cDNA fragment containing gene 73.
XX
KW   Human; secreted protein; treatment; prevention; protein therapy; AIDS;
KW   gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
KW   developmental abnormality; fetal deficiency; blood disorder; leukemia;
KW   immune system disease; autoimmune disease; hepatic disease; lymphoma;
KW   renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
KW   cognitive disorder; prostate disease; skeletal, cardiac, muscle disorder;
KW   pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
KW   arthritis; malignancy; digestive; endocrine; infection; ss.
XX
OS   Homo sapiens.
XX
PN   WO9918208 A1.
XX
PD   15 APR 1999.
XX
PF   01 OCT 1998;   98WO-US20775.
XX
PR   02 OCT 1997;   97US-0060884.
PR   02 OCT 1997;   97US-0060833.
PR   02 OCT 1997;   97US-0060836.
PR   02 OCT 1997;   97US-0060837.
PR   02 OCT 1997;   97US-0060838.
PR   02 OCT 1997;   97US-0060839.
PR   02 OCT 1997;   97US-0060843.
PR   02 OCT 1997;   97US-0060862.
PR   02 OCT 1997;   97US-0060866.
PR   02 OCT 1997;   97US-0060874.
XX
PA   (HUMA ) HUMAN GENOME SCI INC.
XX
PI   Carter KC,  Doan DK,  Endress GA,  Feng P,  Ferrie AM;
PI   Florence KA,  Greene JM,  Janat F,  Lafleur DW,  Ni J;
PI   Rosen CA,  Ruben SM,  Shi Y,  Young P,  Yu G;
XX
DR   WPI; 1999-264022/22.
DR   P-PSDB; AAY07924.
XX
PT   New isolated human genes and the secreted polypeptides they encode
XX
PS   Claim 1a; Page 263-264; 868pp; English.
XX
CC   This invention describes novel isolated human genes and the secreted
CC   proteins they encode. The products of the invention are useful for
CC   prevention, treatment or amelioration medical conditions, e.g. by protein
CC   or gene therapy. Also pathological conditions can be diagnosed by
CC   determining the amount of the new polypeptides in a sample or by
CC   determining the presence of mutations in the new polynucleotides.
CC   Specific uses are described for each of the 101 polynucleotides, based
CC   on which tissues they are most highly expressed in, and include
CC   developing products for the diagnosis or treatment of cancer, tumours,
CC   neurodegenerative disorders, developmental abnormalities and fetal
CC   deficiencies, blood disorders, leukemias, diseases of the immune system,
CC   autoimmune diseases, hepatic and renal disease, lymphomas, inflammation,
CC   allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate
CC   disease, skeletal or cardiac muscle disorders, pulmonary disorders,
CC   transplant rejection, disorders involving osteoclasts such as
CC   osteoporosis, arthritis or malignancies, digestive/endocrine disorders,
CC   infections and AIDS. The human secreted proteins of the invention are
CC   represented in AAY07852 Y07993 and the encoding nucleic acids are
CC   represented in AAX37451 X37552.
XX
SQ   Sequence 868 BP; 272 A; 139 C; 166 G; 291 T; 0 other;


alignment scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2 COPY_66_74 x AAX37523   ..

 Align seq 1/1  to: AAX37523  from: 1  to: 868

       2 TyrPro..GluAspSerSer 7
         ||||||||||||||||||
     543 TATCCTGAGGATTCCTCA 560

seq_name: /cgbin/..prdata/geneseq.geneseqp_NA1999.DAT,AAI77777

seq documentation block:
ID   AAI77777 standard; DNA; 900 BP.
XX
AC   AAI77777;
XX
DT   30-SEP-1997  (first entry)
XX
DE   Fragment of SR Ca2+ release channel coding sequence from IDC patient.
XX
KW   Wild type; mutant; human; sarcoplasmic reticulum calcium release channel;
KW   ischaemic cardiomyopathy; idiopathic dilated cardiomyopathy; IDC;
KW   identity; predisposition; detection; genetic susceptibility; primer;
KW   HindIII site; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
```

```
FT   mutation          230
FT                     /*tag   a
FT                     /note  "Represents mutant residue, G>A"
FT   mutation          626
FT                     /*tag   b
FT                     /note  "Represents mutant residue, G>T"
XX
PN   US5639614-A.
XX
PD   17-JUN-1997.
XX
PF   07-JUN-1995;    95US-0480481.
XX
PR   07-JUN-1995;    95US-0480481.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Lusis A.J,  Philipson K.D,  Sen L;
XX
DR   WPI; 1997-332042/30.
XX
PT   Detecting genetic pre disposition to idiopathic dilated
PT   cardiomyopathy - by detecting mutation(s) in sarcoplasmic reticulum
PT   calcium release channel gene, as a HindIII RFLP
XX
PS   Example 12, Column 27 28, 23pp, English.
XX
CC   This sequence represents a fragment of the mutant human sarcoplasmic
CC   reticulum calcium release channel coding sequence amplified using the
CC   primer sequences given in AAT77775-76.  The mutant sequence was
CC   amplified from an individual suffering from idiopathic dilated
CC   cardiomyopathy (IDC).  The presence of the mutant sequence in a sample
CC   may be used to identify an individual who is genetically predisposed to
CC   IDC. Diagnostic kits for detecting genetic susceptibility to IDC,
CC   comprise a pair of oligonucleotide primers for amplifying a segment of
CC   the SR Ca++ release channel gene, which are homologous to sequences on
CC   either side of the HindIII site at position 777, and a HindIII enzyme for
CC   cleaving DNA amplified by PCR using the primer pair.
XX
SQ   Sequence 900 BP; 299 A; 184 C; 191 G; 226 T; 0 other;


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_66_74 x AAT77777   ..

 Align seq 1/1  to: AAT77777  from: 1  to: 900

       1 LeuTyrProGluAspSer 6
         ||||||||||||||||||
     572 CTGTATCCAGAGGATTCA 589

seq_name: /cds1/ftp-data/geneseq/patseqs_NS1-2/AAT77777

seq_documentation_block:
ID   AAT60094 standard; DNA; 900 BP.
XX
AC   AAT60094;
XX
DT   10-JUN-1997  (first entry)
XX
DE   Human SR calcium ion release channel mutant gene fragment.
XX
KW   Idiopathic dilated cardiomyopathy; IDS; diagnosis; gene therapy;
KW   sarcoplasmic reticulum calcium ion release channel; RFLP;
KW   restriction fragment length polymorphism; ds.
XX
OS   Homo sapiens
XX
```

```
FH   Key               Location/Qualifiers
FT   misc_feature      1
FT                     /*tag   a
FT                     /note  "position 1 corresponds to 8137 of the
FT                     rabbit cDNA homologue"
FT   mutation          230
FT                     /*tag   b
FT                     /note  "position 230 (guanosine in the wild type
FT                     gene) distinguishes IDC"
FT   mutation          626
FT                     /*tag   c
FT                     /note  "position 626 (guanosine in the wild type
FT                     gene) distinguishes IDC"
XX
PN   WO9711089-A1.
XX
PD   27-MAR-1997.
XX
PF   03-SEP-1996;    96WO-US14081.
XX
PR   19-SEP-1995;    95US-0003971.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Lusis A.J,  Philipson K.D,  Sen L;
XX
DR   WPI; 1997-202813/18.
XX
PT   Diagnosis and treatment of idiopathic dilated cardiomyopathy - by
PT   identifying and modulating a mutant sarcoplasmic reticulum calcium
PI   ion release channel gene having point mutation(s)
XX
PS   Example 12; Page 45-47; 61pp; English.
XX
CC   A fragment (AAI60094) of the human sarcoplasmic reticulum (SR)
CC   calcium ion release channel gene is isolated from patients
CC   suffering from idiopathic dilated cardiomyopathy (IDC) using PCR
CC   primers (see also AAT60092-93) that correspond to nucleotides
CC   9007-9036 and 8137 8166 of the rabbit SR Ca2+ release channel gene.
CC   Its sequence differs at 2 points from that of a DNA fragment
CC   amplified, using the same primers, from normal subjects or patients
CC   suffering from ischaemic cardiomyopathy.  The 2 point mutations can
CC   be utilised in methods, e.g. RFLP analysis, for testing a genetic
CC   predisposition to IDC.
XX
SQ   Sequence 900 BP; 299 A; 184 C; 191 G; 226 T; 0 other;


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_66_74 x AAI60094   ..

 Align seq 1/1  to: AAI60094  from: 1  to: 900

       1 LeuTyrProGluAspSer 6
         ||||||||||||||||||
     572 CTGTATCCAGAGGATTCA 589

seq_name: /cds1/ftp-data/geneseq/patseqs_NS1-2/AAI60094

seq_documentation_block:
ID   AAA78484 standard; DNA; 968 BP.
XX
AC   AAA78484;
XX
DT   27-NOV-2000  (first entry)
XX
DE   Plant SEF polynucleotide sequence SEQ List 1 NO:17e.
XX
```

KW Plant; corn; Arabidopsis thaliana; sequence determined DNA fragment;
KW SDF; genetic mapping; identification; promoter; structural gene; UTR;
KW untranslated region; expression control; ds.
XX
OS Plant.
XX
PN WO200040695-A2.
XX
PD 13-JUL-2000.
XX
PF 07-JAN-2000; 2000WO-US00466.
XX
PR 08-JAN-1999; 99US-0115293.
XX
PA (CERE) CERES INC.
XX
PI Alexandrov N, Brover V, Chen X, Subramanian G, Troukhan ME;
PI Zheng L;
XX
DR WPI; 2000-465970/40.
XX
PT New corn plant and Arabidopsis thaliana sequence determined DNA
PT fragments, useful for expressing gene products and for controlling
PT expression of a target gene -
XX
PS Claim 1; Page 407; 674pp; English.
XX
CC The present invention describes polynucleotides, such as complete cDNA
CC sequences and/or sequences of genomic DNA encompassing complete genes,
CC portions of genes, and/or intergenic regions, collectively referred to
CC as sequence-determined DNA fragments (SDFs), from corn plants and
CC Arabidopsis thaliana. The SDFs are promoters, structural genes,
CC untranslated regions (UTRs), or 3' termination sequences. They can be
CC used for expressing a gene product and controlling expression of a
CC target gene, either as a promoter, a structural gene, an UTR or as a
CC 3' termination sequence. They are also useful as tools for genetic
CC mapping, and identification of a particular individual plant or for
CC clustering a group pf plants with a common trait. AAA78433 to AAA78630
CC and AAB24605 to AAB25099 represent the specifically claimed
CC polynucleotide sequences and polypeptides encoded by them given in the
CC present invention.
XX
SQ Sequence 968 BP; 288 A; 196 C; 225 G; 259 T; 0 other;

alignment scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2 (COPY 66_74 x AAA78484/rev ..

Align seq 1/1 to reverse of: AAA78484 from: 1 to: 968

2 TyrProGluAspSerSer 7
||||||||||||||||||
646 TATCCTGAAGATTCGTCA 629

```
RM of: US-09-126-945B-2_COPY_66_74 *** Issued Patents_NA *
Date: Sep 14, 2001  9:53 PM

About: Results were produced by the [illegible]
       Copyright [illegible]

Command line parameters:
  -MODEL frame_p2n model -DEV xlp
  -Q [illegible]
  -DB Issued_Patents_NA -QFMT -fastap -SUFFIX olip2n.rni
  -GAPOP 4.500 -GAPEXT 0.050 -MINMATCH 0.100 -LOOPCL 0.000
  -LOOPEXT 0.000 -gGAPOP 4.500 -gGAPEXT 0.050 -XGAPOP 60.000
  -XGAPEXT 60.000 -FGAPOP 6.000 -FGAPEXT 7.000 -YGAPOP 60.000
  -YGAPEXT 60.000 -DELOP 6.000 -DELEXT 7.000 -SIART 1
  -MATRIX oligo -TRANS human40.cdi -LIST 45 -f=0 ALIGN 200
  -THR_SCORE quality -THR_MIN 1 -ALIGN 15 -MODE LOCAL -OUTFMT pir
  -NORM ext -MINLEN 0 -MAXLEN 2000000000
  -USER US09126945_COGNI_1_91 -NCPU 6 -DCPU 3 LONGLOG1 -NO_XLPXY
  -WAIT -THREADS 1

Search information block:
Query: US-09-126-945B-2_COPY_66_74
Query length: 9
Database: Issued_Patents_NA:*
Database sequences: 324599
Database length: 94655562
Search time (sec): 322.720000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000


score_list:
Sequence                Strd Orig    ZScore   EScore Len  ! Documentation  ..
/cgn2_6/ptodata/2/ina/5A_COMB seq:US-08-373-284A-10 +     6.00   113.19   2.21      21  !
/cgn2_6/ptodata/2/ina/5A_COMB seq:US-08-544-215-5 +     6.00   101.10   28.56      142  !
/cgn2_6/ptodata/2/ina/5A_COMB seq:US-08-480-481-6 +     6.00   118.38   67.52      600  !
/cgn2_6/ptodata/2/ina/5A_COMB seq:US-08-359-665-1 +     6.00   117.48   66.76     1061  !
/cgn2_6/ptodata/2/ina/5A_COMB seq:US-08-420-193-3 +     6.00   117.31   65.71     1050  !
/cgn2_6/ptodata/2/ina/5A_COMB seq:US-09-045-193-1 +     6.00   117.31   65.71     1050  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-946-497-3 +     6.00   117.07   66.36     1062  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-483-522-3 +     6.00   117.07   66.36     1062  !
/cgn2_6/ptodata/2/ina/5B_COMB seq:US-08-478-882-3 +     6.00   117.07   66.36     1062  !
/cgn2_6/ptodata/2/ina/5A_COMB seq:US-08-861-915-2 +     6.00   114.98   99.63     1680  !
/cgn2_6/ptodata/2/ina/5A_COMB seq:US-08-861-915-11 +     6.00   111.61   137.51     2307  !
/cgn2_6/ptodata/2/ina/5A_COMB seq:US-08-861-845-1 +     6.00   105.43   167.86     3111  !
/cgn2_6/ptodata/2/ina/5A_COMB seq:US-07-738-042-1 +     6.00   102.27   31.47      256  !  Da
/cgn2_6/ptodata/2/ina/5A_COMB seq:US-08-907-223-121 +     6.00   101.29   211.84     3500  !
/cgn2_6/ptodata/2/ina/5A_COMB seq:US-08-621-660A-1 +     6.00   101.29   211.84     3500  !
/cgn2_6/ptodata/2/ina/5B_COMB seq:US-08-702-105A-21 +     6.00   101.16   508.26      135  !
/cgn2_6/ptodata/2/ina/5B_COMB seq:US-08-702-105A-22 +     6.00   101.16   508.26      135  !
/cgn2_6/ptodata/2/ina/5A_COMB seq:US-08-702-105A-24 +     6.00   101.16   508.26      135  !
/cgn2_6/ptodata/2/ina/5A_COMB seq:US-08-702-105A-26 +     6.00   101.16   508.26      135  !
/cgn2_6/ptodata/2/ina/Backfile1 seq:5427005-1 +     6.00   100.75   116.34     1600  ! Data
/cgn2_6/ptodata/2/ina/5A_COMB seq:US-09-040-614-14 +     6.00   100.63   115.45      551  !
/cgn2_6/ptodata/2/ina/Backfile1 seq:5175117-1 +     6.00   100.63   115.45      454  ! Data
/cgn2_6/ptodata/2/ina/5B_COMB seq:US-08-908-416-225 +     6.00   101.61   481.65      538  !
/cgn2_6/ptodata/2/ina/5A_COMB seq:US-09-073-660-21 +     6.00   101.44   502.58      517  !
/cgn2_6/ptodata/2/ina/5B_COMB seq:US-08-425-278-13 +     6.00   101.44   502.58      517  !
/cgn2_6/ptodata/2/ina/5A_COMB seq:US-08-257-784A-12 +     5.00   101.16   510.39      570  !
/cgn2_6/ptodata/2/ina/5B_COMB seq:US-09-499-066A-1 +     5.00   101.16   510.49      570  !
/cgn2_6/ptodata/2/ina/5A_COMB seq:US-08-480-672A-1 +     5.00   101.16   510.49      570  !
/cgn2_6/ptodata/2/ina/5B_COMB seq:US-08-702-105A-28 +     5.00   100.99   521.97      585  !
/cgn2_6/ptodata/2/ina/5B_COMB seq:US-08-702-105A-30 +     5.00   100.99   521.97      585  !
/cgn2_6/ptodata/2/ina/5A_COMB seq:US-09-702-105A-28 +     5.00   100.99   521.97      585  !
/cgn2_6/ptodata/2/ina/5A_COMB seq:US-08-702-105A-30 +     5.00   100.99   521.97      585  !
/cgn2_6/ptodata/2/ina/5A_COMB seq:US-09-342-921-9 +     5.00   100.96   522.51      597  !
/cgn2_6/ptodata/2/ina/5A_COMB seq:US-09-542-120-123 +     5.00   100.85   523.14      642  !
/cgn2_6/ptodata/2/ina/5B_COMB seq:US-08-908-416-760 +     5.00   99.50   631.19      729  !
/cgn2_6/ptodata/2/ina/5A_COMB seq:US-09-383-664-10 +     5.00   99.07   666.92      777  !
/cgn2_6/ptodata/2/ina/5B_COMB seq:US-09-09-924-908-1 +     5.00   98.96   692.46      798  !
/cgn2_6/ptodata/2/ina/Backfile1 seq:5217084-1 +     5.00   98.65   711.55      828  ! Da
/cgn2_6/ptodata/2/ina/5A_COMB seq:US-09-880-503-20 +     5.00   98.42   733.85      868  !
/cgn2_6/ptodata/2/ina/5B_COMB seq:US-09-260-634B-5 +     5.00   97.39   827.83      998  !
/cgn2_6/ptodata/2/ina/5A_COMB seq:US-08-202-386-20 +     5.00   96.89   882.69     1075  !
/cgn2_6/ptodata/2/ina/5A_COMB seq:US-08-474-068-2 +     5.00   96.89   882.69     1075  !
/cgn2_6/ptodata/2/ina/5B_COMB seq:US-09-908-727-2 +     5.00   96.86   886.24     1080  !
/cgn2_6/ptodata/2/ina/5A_COMB seq:US-09-265-680A-2 +     5.00   96.86   886.24     1080  !

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB seq:US-08-373-284A-10

seq_documentation_block:
; Sequence 10, Application US/08473284A
; Patent No. 5830646
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  DIAGNOSTIC METHOD
;    NUMBER OF SEQUENCES:  14
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  MS-DOS 7
;      SOFTWARE:  WordPad for Windows 95
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/473,284A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9215498.8
;      FILING DATE:  21-JUL-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9224386.4
;      FILING DATE:  20-NOV-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9226165.0
;      FILING DATE:  16-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/GB93/01520
;      FILING DATE:  20-JUL-1993
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon & Vanderhye P.C.
;      STREET:  1100 No. 90[illegible]th Glebe Road, 8th Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22201-4714
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Leonard C Mitchard
;      REGISTRATION NUMBER:  29009
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)-816-4000
;      TELEFAX:  (703)-816-4100
;      TELEX:  [illegible] NIXN UR
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  21 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-373-284A-10


Alignment_scores:
            Quality:    6.00              Length:      6
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment_block:
US-09-126-945B-2_COPY_66_74 x US-08-373-284A-10  ..

Align_seq 1/1  1 .. US-09-373-284A-10  1 .. 21
```

```
        4 ProGlnAspSerSerTrp 8
          ||||||||||||||||||
        4 CCAGAGGACAGTTCCTGG 21

Seq name: /cgn2_6/ptodata/1/iss_2/A_5MB.seq:US-08-564-225-5

Seq documentation block:
; Sequence 5, Application US/08564225
; Patent No. 6010865
;  GENERAL INFORMATION:
;     APPLICANT:  Ponta, Helmut
;     APPLICANT:  Heider, Karl Heinz
;     APPLICANT:  Herrlich, Peter
;     APPLICANT:  Rals, Steven I.
;     APPLICANT:  Hall, Peter
;     TITLE OF INVENTION:  Processes For Diagnosing And Analysing Tumours.
;     NUMBER OF SEQUENCES:  12
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
;       ADDRESSEE:  Dunner, L.L.P.
;       STREET:  1300 I Street, N.W., Suite 700
;       CITY:  Washington
;       STATE:  D.C.
;       COUNTRY:  USA
;       ZIP:  20005-3315
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/564,225
;       FILING DATE:  04 JUN 1996
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  GR P 43 14 787.2
;       FILING DATE:  28 APR 1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  GR P 43 21 944.6
;       FILING DATE:  02 JUL 1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  GR P 43 20 624.9
;       FILING DATE:  22 JUN 1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  GR P 43 20 624.7
;       FILING DATE:  22 JUN 1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Barker, Paul
;       REGISTRATION NUMBER:  32,013
;       REFERENCE/DOCKET NUMBER:  05814.0003-00000
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 408-4000
;       TELEFAX:  (202) 408-4400
;   INFORMATION FOR SEQ ID NO:  5:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  142 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  3..141
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  1..142
US-08-564-225-5


Alignment scores:
            Quality:     6.00              Length:       6
              Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000
```

```
Alignment block:
 US-09-126-945B-2-COPY-66-74 x US-08-564-225-5

 Alignment:  1 to 142  of  142 in US-08-564-225-5  from:  1  to:  142

        4 ProGlnAspSerSerTrp 8
          |||| ||||||||||||||
       51 CCAGAGGACAGTTCCTGG 68

Seq name: /cgn2_6/ptodata/1/iss_2/A_5MB.seq:US-08-480-481-6

Seq documentation block:
; Sequence 6, Application US/08480481
; Patent No. 5639614
;  GENERAL INFORMATION:
;     APPLICANT:  Seo, Duyl
;     APPLICANT:  Phillipson, Kenneth D.
;     APPLICANT:  Lusis, Aldons J.
;     TITLE OF INVENTION:  GENE MUTATION IN PATIENTS WITH IDIOPATHIC DILATED CARD
;     NUMBER OF SEQUENCES:  6
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Knobbe, Martens, Olson & Bear
;       STREET:  620 Newport Center Drive  16th Floor
;       CITY:  Newport Beach
;       STATE:  CA
;       COUNTRY:  U.S.A.
;       ZIP:  92660
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ Version 1.5
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/480,481
;       FILING DATE:  07-JUN-1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Altman, Daniel E.
;       REGISTRATION NUMBER:  34,115
;       REFERENCE/DOCKET NUMBER:  UCLA012.001A
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  714 760 0404
;       TELEFAX:  714 760 9502
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  900 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  both
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  Genomic DNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     FRAGMENT TYPE:
;     ORIGINAL SOURCE:
;     FEATURE:
;       NAME/KEY:  Other
;       LOCATION:  1...1
;       OTHER INFORMATION:  Position 1 corresponds to
;       OTHER INFORMATION:  8137 of the rabbit cDNA homolog
;       NAME/KEY:  Other
;       LOCATION:  230...230
;       OTHER INFORMATION:  Position distinguishes idiopathic
;       OTHER INFORMATION:  dilated cardiomyopathy
;       NAME/KEY:  Other
;       LOCATION:  626...626
;       OTHER INFORMATION:  Position distinguishes idiopathic
;       OTHER INFORMATION:  dilated cardiomyopathy
US-08-480-481-6
```

```
alignment_scores:
            Quality:     6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_66_74 x US-08-480-481-6    ..

 Align seq 1/1  to: US-08-480-481-6  from: 1  to: 900

        1 LeuTyrProGluAspSer 6
          ||||||||||||||||||
      572 CTGTATCCAGAGGATTCA 589

seq_name: /tmp2/pir/dir1/fos_SR_2MB_seq/US-08-480-481-6

seq_documentation_block:
; Sequence 1, Application US/08359850
; Patent No. 5951982
;  GENERAL INFORMATION:
;    APPLICANT:  Zoller, Margot
;    APPLICANT:  Herrlich, Peter
;    APPLICANT:  Ponta, Helmut
;    TITLE OF INVENTION:  Use of Antibody Containing Preparations
;    TITLE OF INVENTION:  for Immunosuppression
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein and Fox
;      STREET:  1225 Connecticut Avenue
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/359,850
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/963,323
;      FILING DATE:
;      APPLICATION NUMBER:  DE P 41 34 982.2
;      FILING DATE:  23-OCT-1991
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 833-7533
;      TELEFAX:  (202) 833-8716
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1014 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
US-08-359-850-1


alignment_scores:
            Quality:     6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_66_74 x US-08-359-850-1    ..

 Align seq 1/1  to: US-08-359-850-1  from: 1  to: 1014

        3 ProGluAspSerSerTrp 8
          ||||||||||||||||||
      537 CCAGAGGACAGTTCCTGG 554

seq_name: /tmp2/pir/dir1/fos_SR_2MB_seq/US-08-480-481-3

seq_documentation_block:
; Sequence 3, Application US/08480481
; Patent No. 5639614
;  GENERAL INFORMATION:
;    APPLICANT:  Seo, Luyi
;    APPLICANT:  Philipson, Kenneth D.
;    APPLICANT:  Lusis, Aldons J.
;    TITLE OF INVENTION:  GENE MUTATION IN PATIENTS WITH IDIOPATHIC DILATED CARDI
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  620 Newport Center Drive  16th Floor
;      CITY:  Newport Beach
;      STATE:  CA
;      COUNTRY:  U.S.A.
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,481
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Altman, Daniel E
;      REGISTRATION NUMBER:  34,115
;      REFERENCE/DOCKET NUMBER:  DCLA012.001A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  714-760-0404
;      TELEFAX:  714-760-9502
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1050 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
;    FEATURE:
;      NAME/KEY:  Other
;      LOCATION:  1...1
;      OTHER INFORMATION:  Position 1 corresponds to
;      OTHER INFORMATION:  7987 of the rabbit cDNA homolog
US-08-480-481-3


alignment_scores:
            Quality:     6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_66_74 x US-08-480-481-3    ..

 Align seq 1/1  to: US-08-480-481-3  from: 1  to: 1050

        1 LeuTyrProGluAspSer 6
          ||||||||||||||||||
      722 CTGTATCCAGAGGATTCA 739
```

```
Seq name: /tmp/pbrian/fisa/A_tMP.seq  Len: 1050

Seq documentation block:
; Sequence 4, Application US/08480481
; Patent No. 5649614
;  GENERAL INFORMATION:
;     APPLICANT:  Sen, Luy:
;     APPLICANT:  Philipson, Kenneth D.
;     APPLICANT:  Lusis, Aldons J.
;     TITLE OF INVENTION:  GENE MUTATION IN PATIENTS WITH HYPERTHYROIDISM
;     NUMBER OF SEQUENCES:  6
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Knobbe, Martens, Olson & Bear
;       STREET:  620 Newport Center Drive  16th Floor
;       CITY:  Newport Beach
;       STATE:  CA
;       COUNTRY:  U.S.A.
;       ZIP:  92660
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ Version 1.5
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/480,481
;       FILING DATE:  07-JUN-1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Altman, Daniel E
;       REGISTRATION NUMBER:  34,115
;       REFERENCE/DOCKET NUMBER:  [illegible]CA
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  714 760-0404
;       TELEFAX:  714 760 9502
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1050 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  Genomic DNA
;     HYPOTHETICAL:  NO
;     ANTI SENSE:  No
;     FRAGMENT TYPE:
;     ORIGINAL SOURCE:
;     FEATURE:
;       NAME/KEY:  other
;       LOCATION:  380...380
;       OTHER INFORMATION:  Position distinguishes idiopathic
;       OTHER INFORMATION:  dilated cardiomyopathy
;       NAME/KEY:  other
;       LOCATION:  776...776
;       OTHER INFORMATION:  Position distinguishes idiopathic
;       OTHER INFORMATION:  dilated cardiomyopathy
;       NAME/KEY:  other
;       LOCATION:  1...1
;       OTHER INFORMATION:  Position 1 corresponds to
;       OTHER INFORMATION:  7987 of the rabbit cDNA homolog
US 08 480 481 4


alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US 09 126 945B 2 COPY 66_74 x US 08 480 481 4    ..

 Align Set 1/1:  to: US 08 480 481 4  from: 1  to: 1050

        1 LeuTyrProGluAspSer 6
          ||||:|||||||||||
      722 CTGTATCCAGAGGATTCA 739


Seq name: /tmp/pbrian/fisa/A_tMP.seq US 07 946 497 4

Seq documentation block:
; Sequence 4, Application US/07946497
; Patent No. 5506119
;  GENERAL INFORMATION:
;     APPLICANT:  HERRLICH, Peter
;     APPLICANT:  PONTA, Helmut
;     APPLICANT:  GUENTHERT, Ursula
;     APPLICANT:  MATZKU, Siegfried
;     APPLICANT:  WENZEL, Achim
;     TITLE OF INVENTION:  VARIANT CD44 SURFACE PROTEINS, DNA
;     TITLE OF INVENTION:  SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE PROTEINS,
;     TITLE OF INVENTION:  AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY
;     NUMBER OF SEQUENCES:  8
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Foley & Lardner
;       STREET:  3000 K Street, N.W., Suite 500
;       CITY:  Washington, D.C.
;       COUNTRY:  USA
;       ZIP:  20007-5109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/946,497
;       FILING DATE:  19921109
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BENT, Stephen A.
;       REGISTRATION NUMBER:  29,768
;       REFERENCE/DOCKET NUMBER:  16915/145
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202)672-5300
;       TELEFAX:  (202)672-5399
;       TELEX:  904195
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1062 base pairs
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     IMMEDIATE SOURCE:
;       CLONE:  human cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..1062
US 07 946 497 3


alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US 09 126 945B 2 COPY 66_74 x US 07 946 497 3    ..

 Align Set 1/1:  to: US 07 946 497 3  from: 1  to: 1062

        8 ProGluAspSerSerTrp 8
          ||||:|||||:|||||
      580 CCAGAGGACAGTTGGTGG 597
```

```
seq_name: /cgn2_6/ptodata/2/iea/5A_COMB.seq:US-08-483-322-3

seq_documentation_block:
; Sequence 3, Application US/08483322
; Patent No. 5760178
;  GENERAL INFORMATION:
;     APPLICANT:  HERRLICH, Peter
;     APPLICANT:  PONTA, Helmut
;     APPLICANT:  GUENTHERT, Ursula
;     APPLICANT:  MATZKU, Siegfried
;     APPLICANT:  WENZL, Achim
;     TITLE OF INVENTION:  VARIANT CD44 SURFACE PROTEINS, DNA
;     TITLE OF INVENTION:  SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE PROTEINS,
;     TITLE OF INVENTION:  AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY
;     NUMBER OF SEQUENCES:  8
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Foley & Lardner
;       STREET:  3000 K Street, N.W., Suite 500
;       CITY:  Washington, D.C.
;       COUNTRY:  USA
;       ZIP:  20007-5109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/483,322
;       FILING DATE:  07-JUN-1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/946,497
;       FILING DATE:  09-NOV-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BENT, Stephen A.
;       REGISTRATION NUMBER:  29,768
;       REFERENCE/DOCKET NUMBER:  16915/145
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202)672-5300
;       TELEFAX:  (202)672-5399
;       TELEX:  904136
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1062 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     IMMEDIATE SOURCE:
;       CLONE:  human cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..1062
US-08-483-322-3


alignment_scores:
            Quality:     6.00              Length:       6
              Ratio:  1.000                 Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_66_74 x US-08-483-322-3    ..

 Align seq 1/1  to: US-08-483-322-3  from: 1  to: 1062

        3 ProGluAspSerSerTrp 8
          ||||||||||||||||||
      580 CCAGAGGACAGTTCCTGG 597

seq_name: /cgn2_6/ptodata/2/iea/5B_COMB.seq:US-08-483-322-3

seq_documentation_block:
```

```
; Application US/08478882
; Patent No. 5880574
;  GENERAL INFORMATION:
;     APPLICANT:  HERRLICH, Peter
;     APPLICANT:  PONTA, Helmut
;     APPLICANT:  GUENTHERT, Ursula
;     APPLICANT:  MATZKU, Siegfried
;     APPLICANT:  WENZL, Achim
;     TITLE OF INVENTION:  VARIANT CD44 SURFACE PROTEINS, DNA
;     TITLE OF INVENTION:  SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE PROTEINS,
;     TITLE OF INVENTION:  AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY
;     NUMBER OF SEQUENCES:  8
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Foley & Lardner
;       STREET:  3000 K Street, N.W., Suite 500
;       CITY:  Washington, D.C.
;       COUNTRY:  USA
;       ZIP:  20007-5109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/478,882
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/946,497
;       FILING DATE:  19921109
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BENT, Stephen A.
;       REGISTRATION NUMBER:  29,768
;       REFERENCE/DOCKET NUMBER:  16915/145
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202)672-5300
;       TELEFAX:  (202)672-5399
;       TELEX:  904136
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1062 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     IMMEDIATE SOURCE:
;       CLONE:  human cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..1062
US-08-478-882-3


alignment_scores:
            Quality:     6.00              Length:       6
              Ratio:  1.000                 Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_66_74 x US-08-478-882-3    ..

 Align seq 1/1  to: US-08-478-882-3  from: 1  to: 1062

        3 ProGluAspSerSerTrp 8
          ||||||||||||||||||
      580 CCAGAGGACAGTTCCTGG 597

seq_name: /cgn2_6/ptodata/2/iea/6A_COMB.seq:US-08-891-845-3

seq_documentation_block:
; Sequence 3, Application US/08891845
; Patent No. 6096873
;  GENERAL INFORMATION:
```

```
;     APPLICANT:  Schaefer, Gabriele M.
;     APPLICANT:  Sliwkowski, Mark
;     TITLE OF INVENTION:  Gamma-Heregulin
;     NUMBER OF SEQUENCES:  11
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Genentech, Inc.
;       STREET:  460 Point San Bruno Blvd
;       CITY:  South San Francisco
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  94080
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  5.5 inch, 1.44 Mb floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  WinPatin (Genentech)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/891,845
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  60/021640
;       FILING DATE:  07/12/96
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Lee, Wendy M.
;       REGISTRATION NUMBER:  40,378
;       REFERENCE/DOCKET NUMBER:  P1043
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415/225 1994
;       TELEFAX:  415/952 9881
;       TELEX:  910/371 7168
;   INFORMATION FOR SEQ ID NO:  8:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1680 base pairs
;       TYPE:  Nucleic Acid
;       STRANDEDNESS:  Single
;       TOPOLOGY:  Linear
US-08-891-845-8


alignment scores:
            Quality:     6.00                Length:       6
              Ratio:  1.000                   Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_66_74 x US-08-891-845-8    ..

 Align seq 1/1  to: US-08-891-845-8  from: 1  to: 1680

        8 ProGlnAspSerSerTrp 8
          ||||||||||||||||||
    1329 CCAGAAGATTCCTCCTGG 1346

seq name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-891-845-11

seq_documentation block:
; Sequence 11, Application US/08891845
; Patent No. 6096873
;  GENERAL INFORMATION:
;     APPLICANT:  Schaefer, Gabriele M.
;     APPLICANT:  Sliwkowski, Mark
;     TITLE OF INVENTION:  Gamma Heregulin
;     NUMBER OF SEQUENCES:  11
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Genentech, Inc.
;       STREET:  460 Point San Bruno Blvd
;       CITY:  South San Francisco
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  94080
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  5.5 inch, 1.44 Mb floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  WinPatin (Genentech)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/891,845
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  60/021640
;       FILING DATE:  07/12/96
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Lee, Wendy M.
;       REGISTRATION NUMBER:  40,378
;       REFERENCE/DOCKET NUMBER:  P1043
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415/225-1994
;       TELEFAX:  415/952 9881
;       TELEX:  910/371 7168
;   INFORMATION FOR SEQ ID NO:  11:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2387 base pairs
;       TYPE:  Nucleic Acid
;       STRANDEDNESS:  Single
;       TOPOLOGY:  Linear
US-08-891-845-11


alignment scores:
            Quality:     6.00                Length:       6
              Ratio:  1.000                   Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_66_74 x US-08-891-845-11    ..

 Align seq 1/1  to: US-08-891-845-11  from: 1  to: 2387

        8 ProGlnAspSerSerTrp 8
          ||||||||||||||||||
    834 CCAGAAGATTCCTCCTGG 851

seq name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-891-845-11

seq_documentation block:
; Sequence 11, Application US/08891845
; Patent No. 6096873
;  GENERAL INFORMATION:
;     APPLICANT:  Schaefer, Gabriele M.
;     APPLICANT:  Sliwkowski, Mark
;     TITLE OF INVENTION:  Gamma Heregulin
;     NUMBER OF SEQUENCES:  11
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Genentech, Inc.
;       STREET:  460 Point San Bruno Blvd
;       CITY:  South San Francisco
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  94080
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  5.5 inch, 1.44 Mb floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  WinPatin (Genentech)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/891,845
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  60/021640
;       FILING DATE:  07/12/96
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Lee, Wendy M.
;       REGISTRATION NUMBER:  40,378
```

```
;       REFERENCE/DOCKET NUMBER:  P1044
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415/225-1994
;       TELEFAX:  415/952 9881
;       TELEX:  910/371 7168
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3111 base pairs
;       TYPE:  Nucleic Acid
;       STRANDEDNESS:  Single
;       TOPOLOGY:  Linear
US-08-891-845-1


alignment_scores:
             Quality:     6.00              Length:       6
               Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_66_74 x US-08-891-845-1    ..

 Align seq 1/1  to: US-08-891-845-1  from: 1  to: 3111

        3 ProGluAspSerSerTrp 8
          ||||||||||||||||||
     1662 CCAGAACATTCCTCCTGG 1679

seq_name: /tmp2_6/pubdata/2/ia1/5A_COMB/seq_US-08-891-845-1.seq 1

seq_documentation_block:
; Sequence 1, Application US/07738032
; Patent No. 5290686
;  GENERAL INFORMATION:
;     APPLICANT:  KENDALL, Alan P.
;     APPLICANT:  BLACK, Renee
;     APPLICANT:  ROTA, Paul A.
;     TITLE OF INVENTION:  IMPROVED EXPRESSION OF INFLUENZA "A" M2
;     TITLE OF INVENTION:  PROTEIN IN BACULOVIRUS AND USES OF M2 PROTEIN
;     NUMBER OF SEQUENCES:  2
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Townsend and Townsend Khourie and Crew
;       STREET:  Steuart Street Tower, One Market Plaza
;       CITY:  San Francisco
;       STATE:  California
;       COUNTRY:  US
;       ZIP:  94105-1493
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/738,032
;       FILING DATE:  19910731
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Weber, Kenneth A.
;       REGISTRATION NUMBER:  31,677
;       REFERENCE/DOCKET NUMBER:  15280-47
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 543-9600
;       TELEFAX:  (415) 543-5043
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  25 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Influenza A/Aichi/2/68 (H3N2) virus
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  1..25
;       OTHER INFORMATION:  /function= "PCR AMPLIFICATION"
;       OTHER INFORMATION:  /product= "PRIMER"
;       OTHER INFORMATION:  /standard_name= "Primer M2F mRNA segment"
US-07-738-032-1


alignment_scores:
             Quality:     6.00              Length:       6
               Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_66_74 x US-07-738-032-1    ..

 Align seq 1/1  to reverse of: US-07-738-032-1  from: 1  to: 25

        4 GluAspSerSerTrp 8
          ||||||||||||||||
       23 CAAGACTCATCTTGG 9

seq_name: /tmp2_6/pubdata/2/ia1/6A_COMB/seq_US-08-905_223_124

seq_documentation_block:
; Sequence 124, Application US/08905223
; Patent No. 6222029
;  GENERAL INFORMATION:
;     APPLICANT:  Edwards, Joan Baptiste D.
;     APPLICANT:  Duelert, Aymeric
;     APPLICANT:  Laroche, Bruno
;     TITLE OF INVENTION:  5' ESTs FOR SECRETED PROTEINS
;     NUMBER OF SEQUENCES:  503
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Knobbe, Martens, Olson & Bear
;       STREET:  501 West Broadway
;       CITY:  San Diego
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  92101-3505
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy Disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  Win95
;       SOFTWARE:  Word
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  08/905,223
;       FILING DATE:
;       CLASSIFICATION:  536
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Israelsen, Ned A.
;       REGISTRATION NUMBER:  29,655
;       REFERENCE/DOCKET NUMBER:
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (619) 235-8550
;       TELEFAX:  (619) 235-0176
;   INFORMATION FOR SEQ ID NO:  124:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  94 base pairs
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  DOUBLE
;       TOPOLOGY:  LINEAR
;     MOLECULE TYPE:  cDNA
;     ORIGINAL SOURCE:
;       ORGANISM:  Homo Sapiens
;       TISSUE TYPE:  Brain
;     FEATURE:
;       NAME/KEY:  sig_peptide
;       LOCATION:  11..85
;       IDENTIFICATION METHOD:  Von Heijne matrix
;       OTHER INFORMATION:  score 4.5
;       /label= SeqMATRIX  seq:MGIWLVGSIVIIF
```

US 08 905 223 124

alignment scores:
Quality: 5.00 Length: 5
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945b-2-copy_66_74 x US 08 905 223 124 .

Align seq 1/1 to: US 08 905 223 124 from: 1 to: 94

5 AspSerSerTrpAla 9
|||||||||||||||||
4 GATTCATCATGGGCA 17

seq_name: /cgn2_6/ptodata/2/zina/5A_COMB.seq:US 08 621 502A 1

seq documentation block:
; Sequence 1, Application US/08621502A
; Patent No. 5763591
; GENERAL INFORMATION:
; APPLICANT: Toyoda Ph.D., Hiroo
; APPLICANT: Formby Ph.D., Bent
; TITLE OF INVENTION: POLYNUCLEIC ACID SEQUENCES THAT ARE
; TITLE OF INVENTION: FUNCTIONALLY ASSOCIATED WITH THE DEVELOPMENT OF AUTOIMMUNE
; TITLE OF INVENTION: DISEASE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/621,502A
; FILING DATE: 25-MAR-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fujita Esq., Sharon M.
; REGISTRATION NUMBER: 38,459
; REFERENCE/DOCKET NUMBER: P07 34324
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213/622-7700
; TELEFAX: 213/489-4210
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 358 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: No. 5763591 Obese Diabetic Mouse
; TISSUE TYPE: Splenocyte
US 08 621 502A 1

alignment scores:
Quality: 5.00 Length: 5
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945b-2-copy_66_74 x US 08 621 502A 1/rev ..

Align seq 1/1 to reverse of: US 08 621 502A 1 from: 1 to: 358

2 TyrProGluAspSer 6
|||||||||||||||||
442 TATCCTGAAGACTCT 428

```
CM of: US-09-126-945B-2_COPY_66_74 to: EST:*   out_format : pls

Date: Sep 14, 2001  7:47 PM

About: Results were produced by the CenCore software, version 1.5.
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
  -MODEL framer_p2n.model  -DEV xlp
  -Q "/tmp/3_09126_spec_USC96_945_out_11_s_1_32_1_1_1_1_1_qry_prot_list_1_1.seq"
  -DB EST -QFMT fastap -SUFFIX olip2n.rst -GAPOP 4.500
  -GAPEXT 0.050 -MINMATCH 0.100  -EXPPVL 0.000  -EXPEXT 0.000
  -QGAPOP 4.500 -QGAPEXT 0.050 -XGAPOP 60.000 -XGAPEXT 60.000
  -FGAPOP 6.000 -FGAPEXT 7.000 -YGAPOP 60.000 -YGAPEXT 60.000
  -DELOP 6.000 -DELEXT 7.000 -START 1 -MATRIX oligo
  -TRANS human40.cdi -LIST 45 -DOCALIGN 200 -THR_SCORE quality
  -THR_MIN 1 -ALIGN 15 -MODE 1 -OUTPUT pls -NORM yes -MINLEN 0
  -MAXLEN 2000000000 -USER US09126945_gCGNI_1_4645 -NCPU 6
  -ICPU 3 -LONGLOG -NO_XLPXY -WAIT -THREADS 1

Search information block.
Query: US-09-126-945B-2_COPY_66_74
Query length: 9
Database: EST:*
Database sequences: 10228115
Database length: 4314594S4
Search time (sec): 9087.630000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000


score_list:
Sequence              Strd  orig    ZScore    EScore  Len  ! Documentation ..
qb_est23:AI685592      +     9.00   285.11    0.3121   486  ! AI685592 [illegible]
qb_est97:BG179568      +     9.00   179.60    0.6643   961  ! BG179568 [illegible]
qb_qss29:AZ018114      +     7.00   149.82    31.62    152  ! AZ018114 RPCI-23 43407.2V RPCI-
qb_est97:BG155214      +     7.00   146.76    46.84    265  ! BG155214 sab41d10.y1 Gm c1026 S
qb_est67:BB527863      -     7.00   146.11    50.90    298  ! BB527863 BB527863 RIKEN full-le
qb_qss33:AZ861257      -     7.00   145.39    55.85    336  ! AZ861257 ZM01470008 Maize IBFI
qb_est38:AV417115      -     7.00   144.98    58.83    359  ! AV417115 AV417115 Lotus japonic
qb_est87:BF380924      -     7.00   144.79    60.40    371  ! BF380924 IL2-UT0074-260800-140-
qb_est76:BE557568      -     7.00   144.33    64.93    401  ! BE557568 tl07c08.y1 Zebrafish R
qb_est79:BE804384            7.00   144.30    64.17    403  ! BE804384 sr79a12.y1 Gm-c1052 Gl
qb_est38:AV420500      -     7.00   144.13    65.62    415  ! AV420500 AV420500 Lotus japonic
qb_est38:AV416465      -     7.00   144.01    66.66    424  ! AV416465 AV416465 Lotus japonic
qb_est38:AV408717            7.00   143.94    67.25    429  ! AV408717 AV408717 Lotus japonic
qb_est70:BE086197      +     7.00   143.94    67.25    429  ! BE086197 PM3 BT0674 260299 002-
qb_est87:BF441874      -     7.00   143.52    70.99    461  ! BF441874 258815 MAGZ SPIG Sus s
qb_qss8:AQ166600       -     7.00   143.10    74.88    495  ! AQ166600 BS_3146_A1_C06_MR CIT
qb_est101:BG509572     -     7.00   143.03    75.56    501  ! BG509572 sad19a02.y1 Gm-c1074 G
qb_est24:AI731451      +     7.00   142.68    79.05    532  ! AI731451 BNLGHi9330 six day old
qb_est69:BE032968      -     7.00   142.53    80.60    546  ! BE032968 133341 MAGZ SPIG Sus s
qb_est84:BF158002      -     7.00   142.39    82.04    559  ! BF158002 1127d02.y1 Soyano Kawa
qb_qss25:AZ352526      -     7.00   141.85    87.92    613  ! AZ352526 IM0090H20R Mouse 10kI
qb_est87:BF425059      -     7.00   141.81    88.35    617  ! BF425059 su53q07.y1 Gm-c1069 Gl
qb_est76:BE612297      -     7.00   141.65    90.13    634  ! BE612297 sr36d07.y1 Gm-c1049 Gl
qb_est25:AI861513      +     7.00   141.57    91.13    643  ! AI861513 614014H04.x1 614 - roo
qb_est45:AW344362      -     7.00   141.21    94.26    672  ! AW344362 tj76t07.y1 Soyano Kawa
qb_est43:AW184702      -     7.00   140.96    98.48    713  ! AW184702 tj17b05.y1 Zebrafish s
qb_est43:AW133619      -     7.00   139.93   112.38    850  ! AW133619 tj09c05.y1 Soyano Kawa
qb_qss3:CNS0229C       +     7.00   139.83   113.96    866  ! AL177897 Tetraodon nigroviridis
qb_est100:BG387670     -     7.00   139.43   119.84    926  ! BG387670 602412526F1 NIH_MGC_92
qb_est76:BE563288      +     7.00   139.39   120.52    933  ! BE563288 601335410F1 NIH_MGC_38
qb_est80:BE871340      +     7.00   139.31   121.68    945  ! BE871340 601449457F1 NIH_MGC_67
qb_qss3:CNS0257K       -     7.00   138.85   129.15   1023  ! AL181721 Tetraodon nigroviridi
qb_est50:AW678955      -     6.00   131.95   312.75    104  ! AW678955 WS1_21_B05.b1_A002 Wat
qb_est70:BE080078            6.00   131.52   330.65    112  ! BE080078 RC4 BT062m 270600 013
qb_est78:BE766029      +     6.00   130.92   356.91    124  ! BE766029 IL3-NT0104-020600-158-
qb_est91:BF735753      +     6.00   129.82   429.26    168  ! BF735753 CM1_AN0080 081100 546
qb_est70:BE107883      +     6.00   128.91   462.27    175  ! BE107883 UI-R-CA0-awx-h-10-0-UI
qb_est56:BB074509      -     6.00   128.84   464.22    177  ! BB074509 BB074509 RIKEN full le
qb_est70:BE115581      +     6.00   128.68   476.09    182  ! BE115581 UI-R-BJ1-avw-e-10-0-UI
qb_est110:WI0841       -     6.00   128.68   476.09    182  ! WI0841 mellhe5.r1 Soares mel
qb_est95:BG004712      -     6.00   128.30   499.47    194  ! BG004712 MR3 3N0185 211100 0
qb_est94:BF926786      -     6.00   128.18   507.18    198  ! BF926786 IL5_NT0228 041200 3
qb_est9:AA612261       -     6.00   127.84   530.09    210  ! AA612261 vo04103.r1 Stratage
qb_est27:AI392294      -     6.00   127.84   530.09    210  ! AI392294 AI392294 dkt2426 3a
qb_est43:AW144107      -     6.00   127.76   535.77    213  ! AW144107 EST294403 Normalize

seq_name: qb_est23:AI685592

seq_documentation_block:
LOCUS       AI685592          486 bp    mRNA             EST       15-DEC-2000
DEFINITION  wz26c07.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2251596 3'
            similar to contains PIR5:I3 MSR1 repetitive element ;, mRNA
            sequence.
ACCESSION   AI685592
VERSION     AI685592.1  GI:4896886
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 486)
  AUTHORS   [illegible]
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgaps-r@mail.nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov/image/image.html
            Insert Length: 1143  Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 480.
FEATURES             Location/Qualifiers
     source          1..486
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2251596"
                     /clone_lib="NCI_CGAP_Pr28"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT7T3 Pac (Pharmacia)
                     with a modified polylinker; Plasmid DNA from the
                     normalized library NCI_CGAP_Pr28 was prepared, and ss
                     circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (clonelibs
                     985e98-986759, 1101192-1101959, and 1217028-1220621)."
BASE COUNT        72 a    140 c    110 g     64 t
ORIGIN


alignment_scores:
              quality:     9.00              Length:      9
                Ratio:    1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_66_74 x AI685592   ..

 Align Set 1/1  to: AI685592  from: 1  to: 486

       ! LeuIlePro11eAspSerSerIrpAla 9
         |||||||||||||||||||||||||||
```

```
     356 CTGTACCCTGAGGACAGCAGCTGGGCA 382

seq name: qb_est97:BG179568

seq documentation block:
LOCUS       BG179568       991 bp    mRNA            EST       06 FEB 2001
DEFINITION  602428182F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4429413 5',
            mRNA sequence.
ACCESSION   BG179568
VERSION     BG179568.1  GI:12686271
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 991)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbcr@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10181  row: d  column: 22
            High quality sequence stop: 655.
FEATURES             Location/Qualifiers
     source          1..991
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4429413"
                     /clone_lib="NIH_MGC_91"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (plaque resistant)"
                     /note="Organ: prostate; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo dT primed.
                     Average insert size 1.4 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH MGC Library."
BASE COUNT      190 a    351 c    305 g    145 t
ORIGIN


alignment scores:
            Quality:     9.00            Length:       9
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-9458-2 COPY 66_74 x BG179568    ..

 Align seq 1/1   to: BG179568  from: 1  to: 991

        1 LeuTyrProGluAspSerSerTrpAla 9
          |||||||||||||||||||||||||||
       99 CTGTACCCTGAGGACAGCAGCTGGGCA 125

seq name: qb_gss20:AZ018114

seq documentation block:
LOCUS       AZ018114       157 bp    DNA             GSS       25 FEB 2000
DEFINITION  RPCI-23-334C7.TV RPCI-23 Mus musculus genomic clone RPCI-23-334C7,
            DNA sequence.
ACCESSION   AZ018114
VERSION     AZ018114.1  GI:7093498
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 157)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinrect
            ,B., Levins,M., Mequanm,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     other_GSSs: RPCI-23-334C7.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.chori.org/orderinfoframe.htm)
            or from Research Genetics (info@resgen.com). BAC end pairs:
            http://www.tigr.org/tdb/mgc/mouse_bac_ends/mouse_bac_end_intro.html
            Plate: 334  row: C  column: 7
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..157
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-334C7"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6 mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).  "
BASE COUNT       36 a     41 c     17 g     63 t
ORIGIN


alignment scores:
            Quality:     7.00            Length:       7
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-9458-2 COPY 66_74 x AZ018114    ..

 Align seq 1/1   to: AZ018114  from: 1  to: 157

       1 LeuTyrProGluAspSerSer 7
         |||||||||||||||||||||
      45 CTATACCCAGAAGATTCTTCA 65

seq name: qb_est97:BG155214

seq documentation block:
LOCUS       BG155214       265 bp    mRNA            EST       06 FEB 2001
DEFINITION  sab41d11.y1 Gm-c1026 Glycine max cDNA clone GENOME SYSTEMS Gm-NbS
            1b Gm-c1026 4028 5', similar to TR:O65808 O65808 MAGNESIUM
            CHELATASE SUBUNIT., mRNA sequence.
ACCESSION   BG155214
VERSION     BG155214.1  GI:12688878
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
```

```
REFERENCE   1  (bases 1 to 265)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Cryderman,D., Shoemaker
            ,A., Bolla,R., Marra,M., Hillier,L., Kucaba,T., Martin,J., Rock,C.,
            Wylie,T., Underwood,K., Stephoe,M., Theising,B., Allen,M., Powers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schrick
            ,R., Ritter,E., Kohn,S., Shin,T., Yokoyama,Y., Theologis,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information,
            call: (800) 430-0030 or (314) 427 3222 FAX:(888) 919 3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            Possible reversed clone; similarity on wrong strand.
FEATURES             Location/Qualifiers
     source          1..265
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="SPENCER SYSTEM CLONE ID: Gm_c1026_A02"
                     /clone_lib="Gm_c1026"
                     /tissue_type="Senescing leaves, mature plants, greenhouse
                     grown."
                     /lab_host="DH10B"
                     /note="Vector: pZIP5 Par (Pharmacia); Site_1: EcoRI;
                     Site_2: HindIII; This cDNA library was constructed from
                     mRNA isolated from senescing leave tissue of mature
                     greenhouse grown plants. Complementary DNA was synthesized
                     from mRNA using a 3' anchored poly(dT) primer. EcoRI
                     adapters were ligated to the blunt-ended cDNA fragments
                     followed by  digestion with EcoRI and HindIII. The cDNA
                     fragments were directionally cloned into the EcoRI-HindIII
                     restriction site of the pZIP Par vector. The ligated DNA
                     fragments were transformed into DH10B host cells  (Gibco
                     BRL). This library was constructed K. Shoemaker and J.
                     Erpelding."
BASE COUNT        70 a      64 c      49 g      82 t
ORIGIN


alignment_scores:
            Quality:    7.00            Length:       7
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_66_74 x BG155214    .

 Align seq 1/1  to: BG155214  from: 1  to: 265

       1 LeuTyrProGluAspSerSer 7
         |||||||||||||||||||||
     173 CTCTATCCAGAAGATTCATCT 193

seq_name: gb_est67:BB527863

seq_documentation_block:
LOCUS       BB527863       296 bp    mRNA            EST       28-JUL-2000
DEFINITION  BB527863 RIKEN full-length enriched, 15 days embryo head Mus
            musculus cDNA clone D930043I23 3', mRNA sequence.
ACCESSION   BB527863
VERSION     BB527863.1  GI:9579321
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 296)
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ohsugi,M., Yasumoto,K., Lin,B., Sorimachi,K.,
            Watanabe,S., Hayashizaki,Y. and Genome Exploration Research Group
            (RIKEN Genome Science Initiative Team) and Genome Science Group
            (Genome Network Project Core Group)
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
COMMENT     Contact: Hayashizaki Y
            Genome Exploration Research Group, Genomic Sciences Center, RIKEN
            Yokohama Institute and Genome Science Laboratory, Frontier Research
            System, The Institute of Physical and Chemical Research (RIKEN)
            1-22-2 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: +81-45-503-9222
            Fax: +81-45-503-9216
            Email: yosh@gsc.riken.go.jp
            URL: http://genome.gsc.riken.go.jp/
            Full-length enriched cDNA library was constructed by Dr. Hayashizaki
            Y. and Dr. Shibata K. at RIKEN Genome Science Laboratory and Genome
            Exploration Research Group, Genomic Sciences Center, RIKEN, Yokohama
            Institute. Sequencing and analysis were performed by Genome Science
            Laboratory, Genome Exploration Research Group, Genomic Sciences
            Center, RIKEN, Yokohama Institute and Genome Science Group (Genome
            Network Project Core Group). This work was supported in part by a
            Research Grant for the RIKEN Genome Exploration Research Project
            from the Ministry of Education, Culture, Sports, Science and
            Technology of the Japanese Government.
FEATURES             Location/Qualifiers
     source          1..296
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="D930043I23"
                     /clone_lib="RIKEN full-length enriched, 15 days embryo
                     head"
                     /sex="mixed"
                     /tissue_type="head"
                     /dev_stage="15 days embryo"
                     /lab_host="DH10B"
                     /note="Site_1: SalI, Site_2: BamHI, cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo activated reverse
                     transcriptase and subsequently enriched for full length by
                     cap trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3'], cDNA
                     was cloned into the XhoI and BamHI sites. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I"
BASE COUNT        69 a      94 c      42 g      91 t
ORIGIN
```

```
alignment scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_66_74 x BB527863/rev   ..

 Align seq 1/1  to reverse of: BB527863  from: 1  to: 296

        3 ProGluAspSerSerTrpAla 9
          ||||||||||||||||||||
       66 CCAGAAGACAGTAGTTGGGC 46

Seq name: qb_gss33:AZ861257

Seq documentation block:
LOCUS        AZ861257       335 bp    DNA              GSS       21-FEB-2001
DEFINITION  ZM0167D09K Mouse 10kb plasmid UHGC1M library Mus musculus genomic
            clone UHGC2M0167D09 R, DNA sequence.
ACCESSION   AZ861257
VERSION     AZ861257.1  GI:13057396
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 335)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,R., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert length: 10000  Std Error: 0.00
            Plate: 0167  row: D  column: 09
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 335.
FEATURES             Location/Qualifiers
     source          1..335
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UHGC2M0167D09"
                     /clone_lib="Mouse 10kb plasmid UHGC1M library"
                     /sex="Male"
                     /lab_host="E. coli strain XL10 Gold, tl resistant, F'"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/cmdr/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high molar excess. The
                     adaptored DNA was purified and size selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4742114|gb|AF129072.1), a copy number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically competent E. coli XL10 Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT        75 a    111 c     77 g     72 t
ORIGIN


alignment scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_66_74 x AZ861257/rev   ..

 Align seq 1/1  to reverse of: AZ861257  from: 1  to: 335

        3 ProGluAspSerSerTrpAla 9
          ||||||||||||||||||||
       87 CCAGAAGACAGTAGTTGGGCA 67

Seq name: qb_est38:AV417115

Seq documentation block:
LOCUS        AV417115       359 bp    mRNA             EST       23-MAY-2000
DEFINITION  AV417115 Lotus japonicus young plants (two week old) Lotus
            japonicus cDNA clone MWM149d01_r 5', mRNA sequence.
ACCESSION   AV417115
VERSION     AV417115.1  GI:7746293
KEYWORDS    EST.
SOURCE      lotus japonicus.
  ORGANISM  Lotus japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotus.
REFERENCE   1  (bases 1 to 359)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Generation of 7137 non redundant expressed sequence tags from a
            legume, Lotus japonicus
  JOURNAL   DNA Res. 7 (2), 127-140 (2000)
  MEDLINE   20277479
COMMENT     Contact: Yasukazu Nakamura
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: ynakam@kazusa.or.jp, TEL 81-438-52-3935, http://www.kazusa.or.jp/lotus/
FEATURES             Location/Qualifiers
     source          1..359
                     /organism="Lotus japonicus"
                     /db_xref="taxon:34305"
                     /clone="MWM149d01_r"
                     /clone_lib="Lotus japonicus young plants (two week old)"
                     /dev_stage="young plants (two week old)"
                     /note="Vector: pBluescriptII SK ; Site_1: EcoRI; Site_2:
                     XhoI; isolate Miyakojima MG-20"
BASE COUNT       102 a     65 c     94 g     98 t
ORIGIN


alignment scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_66_74 x AV417115/rev   ..

 Align seq 1/1  to reverse of: AV417115  from: 1  to: 359

        1 LeuTyrProGluAspSerSer 7
          |||||||||||||||||||||
      174 CTCTATCCAGAAGACAGTCA 154
```

```
seq_name: gb_est87.BF380924

seq_documentation_block:
LOCUS       BF380924        371 bp    mRNA            EST       27-NOV-2000
DEFINITION  IL2-UT0074-260800-140-E08 UT0074 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF380924
VERSION     BF380924.1  GI:11370049
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 371)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509 010, Sao Paulo SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11 2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/[illegible]
            260800-140-E08&t3=2000-08-26&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 36
            High quality sequence stop. 256.
FEATURES             Location/Qualifiers
     source          1..371
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="UT0074"
                     /dev_stage="Adult"
                     /note="Organ: uterus_tumor; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 - Ludwig Institute for Cancer Research)
                     profiles into the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      103 a    104 c     66 g     97 t      1 others
ORIGIN


alignment_scores:
           Quality:     7.00             Length:       7
             Ratio:    1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_66_74 x BF380924/rev   ..

 Align seq 1/1  to reverse of: BF380924  from: 1  to: 371

       1 LeuTyrProGluAspSerSer 7
         |||||||||||||||||||||
     353 TTATATCCTGAAGATTCATCA 333

seq_name: gb_est76:BE557568

seq_documentation_block:
LOCUS       BE557568        401 bp    mRNA            EST       30-AUG-2000
DEFINITION  fi07c08.y1 Zebrafish Research Genetics C32 Fin Danio rerio cDNA 5'
            similar to TR:P97855 P97855 RAS-GTPASE ACTIVATING PROTEIN
            SH3 DOMAIN BINDING PROTEIN. ;, mRNA sequence.
ACCESSION   BE557568
VERSION     BE557568.1  GI:9822058
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1  (bases 1 to 401)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Other_ESTs: fi07c08.x1
            Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zfish@watson.wustl.edu
            cDNA Library Preparation: Ning Wu. cDNA Library Arrayed by:
            Research Genetics. DNA Sequencing by: Washington University Genome
            Sequencing Center Clone distribution: Research Genetics web
            address: http://www.resgen.com/
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 211.
FEATURES             Location/Qualifiers
     source          1..401
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone_lib="Zebrafish Research Genetics C32 Fin"
                     /tissue_type="Fin"
                     /lab_host="GeneHogs (HS996, a phage resistant isolate of
                     DH10B)"
                     /note="Vector: pT7T3D-Pac with a modified polylinker;
                     Site_1: EcoRI; Site_2: NotI; 1st strand cDNA was prepared
                     from zebrafish(C32) fin, and was then primed with a Not I
                      oligo(dT) primer. Double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector. Library is non-normalized. Library was
                     constructed by Ning Wu. NOTE:Clones from this library are
                     only available thru Research Genetics (www.resgen.com)."
BASE COUNT      103 a     84 c    118 g     96 t
ORIGIN


alignment_scores:
           Quality:     7.00             Length:       7
             Ratio:    1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_66_74 x BE557568/rev   ..

 Align seq 1/1  to reverse of: BE557568  from: 1  to: 401

       2 TyrProGluAspSerSerTrp 8
         |||||||||||||||||||||
     304 TACCCAGAAGACTCCTCTTGG 284

seq_name: gb_est79:BE804384

seq_documentation_block:
LOCUS       BE804384        403 bp    mRNA            EST       20-SEP-2000
DEFINITION  [illegible] cDNA clone GENOME SYSTEMS CLONE ID:
```

```
            Gm_c1052 1943 5' similar to TR:Q05808 Q05808 MAGNESIUM CHELATASE
            SUBUNIT. ;, mRNA sequence.
ACCESSION   BE804484
VERSION     BE804484.1  GI:10235496
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE   1  (bases 1 to 403)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,R., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Papa,L., Barvey,N., Seburk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430 0030 or (314) 427 4222 FAX:(888) 919 4324 or (314)
            427 4324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            High quality sequence stop: 366.
FEATURES             Location/Qualifiers
     source          1..403
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm_c1052 1943"
                     /clone_lib="Gm_c1052"
                     /tissue_type="whole seedlings of greenhouse grown plants"
                     /dev_stage="1 week old"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:
                     XhoI; The Barosoy NIL was constructed and seed was
                     provided by Dr. J. Specht, University of Nebraska
                     (Shoemaker and Specht, 1995). The cDNA library was
                     constructed from mRNA isolated from whole seedlings of 1
                     week old greenhouse grown plants. Complementary DNA was
                     synthesized from mRNA using a primer consisting of a
                     poly(dT) sequence with a XhoI restriction site and a 3'
                     anchor. EcoRI adapters were ligated to the blunt ended
                     cDNA fragments followed by XhoI digestion. The cDNA
                     fragments were directionally cloned into the EcoRI XhoI
                     restriction site of the pBluescript vector. The ligated
                     cDNA fragments were transformed into DH10B host cells
                     (GibcoBRL). The library was constructed in cooperation
                     with Dr. Paul Keim's laboratory at Northern Arizona
                     University."
BASE COUNT       115 a      75 c     107 g     106 t
ORIGIN


alignment scores:
              Quality:     7.00              Length:       7
                Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2_COPY_66_74 x BE804484/rev    .

 Align seq 1/1  to reverse of: BE804484  from: 1  to: 403

        1 LeuTyrProGluAspSerSer 7
          |||||||||||||||||||||
      .
```

```
      179 CTCTATCAGAAGATTCATCT 159

seq name: gb:est38:AV420500

seq documentation block:
LOCUS       AV420500       415 bp    mRNA             EST       23 MAY 2000
DEFINITION  AV420500 Lotus japonicus young plants (two week old) Lotus
            japonicus cDNA clone MWM182c06_r 5', mRNA sequence
ACCESSION   AV420500
VERSION     AV420500.1  GI:7749677
KEYWORDS    EST.
SOURCE      Lotus japonicus.
  ORGANISM  Lotus japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotus.
REFERENCE   1  (bases 1 to 415)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Generation of 7147 non redundant expressed sequence tags from a
            legume, Lotus japonicus
  JOURNAL   DNA Res. 7 (2), 127-130 (2000)
  MEDLINE   20277479
COMMENT     Contact: Yasukazu Nakamura
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: ynakamu@kazusa.or.jp, URL:http://www.kazusa.or.jp/plant/
FEATURES             Location/Qualifiers
     source          1..415
                     /organism="Lotus japonicus"
                     /db_xref="taxon:34305"
                     /clone="MWM182c06_r"
                     /clone_lib="Lotus japonicus young plants (two week old)"
                     /dev_stage="young plants (two week old)"
                     /note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
                     XhoI; isolate Miyakojima MG-20"
BASE COUNT       127 a      92 c      94 g     102 t
ORIGIN


alignment scores:
              Quality:     7.00              Length:       7
                Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2_COPY_66_74 x AV420500/rev    .

 Align seq 1/1  to reverse of: AV420500  from: 1  to: 415

        1 LeuTyrProGluAspSerSer 7
          |||||||||||||||||||||
      29 CTCTATCAGAAGATTCATCT 9

seq name: gb:est38:AV416465

seq documentation block:
LOCUS       AV416465       424 bp    mRNA             EST       23 MAY 2000
DEFINITION  AV416465 Lotus japonicus young plants (two week old) Lotus
            japonicus cDNA clone MWM127d03_r 5', mRNA sequence
ACCESSION   AV416465
VERSION     AV416465.1  GI:7745644
KEYWORDS    EST.
SOURCE      Lotus japonicus.
  ORGANISM  Lotus japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotus.
REFERENCE   1  (bases 1 to 424)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Generation of 7147 non redundant expressed sequence tags from a
            legume, Lotus japonicus
  JOURNAL   DNA Res. 7 (2), 127-130 (2000)
```

```
  MEDLINE   20277479
COMMENT     Contact: Yasukazu Nakamura
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: yasukazu@kazusa.or.jp PRE http://www.kazusa.or.jp/...
FEATURES             Location/Qualifiers
     source          1..424
                     /organism="Lotus japonicus"
                     /db_xref="taxon:34305"
                     /clone="MWM127d03_r"
                     /clone_lib "Lotus japonicus young plants (two-week old)"
                     /dev_stage "young plants (two-week old)"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI; isolate Miyakojima MG-20"
BASE COUNT      126 a     69 c    116 g    113 t
ORIGIN


alignment_scores:
            Quality:    7.00             Length:       7
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_66_74 x AV416465/rev   ..

 Align seq 1/1  to reverse of: AV416465  from: 1  to: 424

       1 LeuTyrProGluAspSerSer 7
         |||||||||||||||||||||
     352 CTCTATCCAGAAGATTCATCT 332

seq_name: qb_est38:AV408717

seq_documentation_block:
LOCUS       AV408717      429 bp    mRNA            EST       24 MAY 2000
DEFINITION  AV408717 Lotus japonicus young plants (two-week old) Lotus
            japonicus cDNA clone MWL045e07_r 5', mRNA sequence.
ACCESSION   AV408717
VERSION     AV408717.1  GI:7721571
KEYWORDS    EST.
SOURCE      Lotus japonicus.
  ORGANISM  Lotus japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotus.
REFERENCE   1  (bases 1 to 429)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Generation of 7137 non-redundant expressed sequence tags from a
            legume, Lotus japonicus
  JOURNAL   DNA Res. 7 (2), 127-130 (2000)
  MEDLINE   20277479
COMMENT     Contact: Yasukazu Nakamura
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: yasukazu@kazusa.or.jp PRE http://www.kazusa.or.jp/...
FEATURES             Location/Qualifiers
     source          1..429
                     /organism="Lotus japonicus"
                     /db_xref="taxon:34305"
                     /clone="MWL045e07_r"
                     /clone_lib="Lotus japonicus young plants (two-week old)"
                     /dev_stage "young plants (two-week old)"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI; isolate Miyakojima MG-20"
BASE COUNT      117 a     81 c    113 g    118 t
ORIGIN


alignment_scores:
            Quality:    7.00             Length:       7
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_66_74 x AV408717/rev   ..

 Align seq 1/1  to reverse of: AV408717  from: 1  to: 429

       1 LeuTyrProGluAspSerSer 7
         |||||||||||||||||||||
     179 CTCTATCCAGAAGATTCATCT 159

seq_name: qb_est70:BE086197

seq_documentation_block:
LOCUS       BE086197      429 bp    mRNA            EST       12 JUN 2000
DEFINITION  BE086197 2632SH 002 c10 BT0674 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE086197
VERSION     BE086197.1  GI:8476590
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 429)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., de Oliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo SP,
            Brazil
            Tel: +55 11 2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            http://www.ludwig.org.br/cgi-bin/ESTdb2.pl?id=BT0674 260
            200-002-e10st3 2000-02-26xt4-1)
            Seq primer: puc 18 forward
            High quality sequence start: 11
            High quality sequence stop: 428.
FEATURES             Location/Qualifiers
     source          1..429
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone_lib "BT0674"
                     /dev_stage "Adult"
                     /note "Organ: breast; Vector: puc18; Site_1: SacI; Site_2:
                     SmaI; A mini library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 - Ludwig Institute for Cancer Research) profiles
                     into the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT      155 a     80 c     86 g    108 t
ORIGIN


alignment_scores:
            Quality:    7.00             Length:       7
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
```

```
 US-09-126-945B-2-COPY-66_74 x BE086197   ..

 Align seq 1/1  to: BE086197  from: 1  to: 429

        1 LeuTyrProGluAspSerSer 7
          |||||||||||||||||||||
      424 CTTTACCCAGAGGACAGCAGC 444

Seq name: gb:est87:BF441874

Seq documentation block:
LOCUS        BF441874       461 bp    mRNA             EST       01-DEC-2000
DEFINITION   258319 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION    BF441874
VERSION      BF441874.1  GI:11501966
KEYWORDS     EST.
SOURCE       pig.
  ORGANISM   Sus scrofa
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE    1  (bases 1 to 461)
  AUTHORS    Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
             Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W
             and Keele,J.W.
  TITLE      Design and use of two pooled tissue normalized cDNA libraries for
             EST discovery in swine
  JOURNAL    Unpublished (2000)
COMMENT      Contact: Smith TPL
             USDA, ARS, US Meat Animal Research Center
             PO Box 166, Clay Center, NE 68933-0166, USA
             Tel: 402 762 4366
             Fax: 402 762 4390
             Email: smith@email.marc.usda.gov
             Single pass sequencing. Bases called and alt_trimmed with phred
             v0.980904.e. Vector identified by cross_match with the -minscore 18
             and -minmatch 12 options.
             PCR Primers
             FORWARD: AGGAAACAGCTATGACCAT
             BACKWARD: GTTTTCCCAGTCACGACG
             Plate: 87  row: P  column: 9
             Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1..461
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 2PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV_SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from testis, ovary,
                     endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT       76 a    149 c    129 g    107 t
ORIGIN


Alignment scores:
            Quality:     7.00            Length:       7
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-126-945B-2-COPY-66_74 x BF441874   ..

 Align seq 1/1  to: BF441874  from: 1  to: 461

        1 LeuTyrProGluAspSerSer 7
          |||||||||||||||||||||
      809 CTTTACCCAGAAGACTCCAGT 829
```

OM of: US-09-126-945B-2_copy_75_83.rni [ ] Run on Monday, MA * [ ] pt

Date: Sep 14, 2001 9:53 PM

About: Results were produced by the [illegible]
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL frames_p2s model -DEV olp
-Q /[illegible]/US-09-126-945B-2_copy_75_83.rni_[1]/app_query.fasta_1.1267
-DB Issued_Patents_NA -OFMT fastap -SUFFIX olip2n.rni
-GAPOP 4.500 -GAPEXT 0.050 -MINMATCH 0.100 -[illegible]
-[illegible]PEXT 0.000 -QGAPOP 4.500 -QGAPEXT 0.050 -XGAPOP 60.000
-XGAPEXT 60.000 -FGAPOP 6.000 -FGAPEXT 7.000 -YGAPOP 60.000
-YGAPEXT 60.000 -DELOP 6.000 -DELEXT 7.000 -START 1
-MATRIX oligo -TRANS human40.cdi -LIST 45 -[illegible]ALIGN 200
-THR_SCORE quality -THR_MIN 1 -ALIGN 15 -MODE [illegible]OCAL -OUTFMT pir
-NORM ext -MINLEN 0 -MAXLEN 2000000000
-USER US09126945_[illegible]CONT_1_91 -NCPU 6 -[illegible]CPU 3 -[illegible]ONSOLE No XLIBXY
-WAIT -THREADS 1

Search information block:
Query: US-09-126-945B-2_COPY_75_83
Query length: 9
Database: Issued_Patents_NA:*
Database sequences: 324599
Database length: 94655562
Search time (sec): 322.720000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	!	Documentation
/[illegible]/2/ina/5B_COMB.seq:US-08-776-971-46	+	6.00	108.84	196.56	246	!	
/[illegible]/2/ina/5B_COMB.seq:US-08-776-971-1[illegible]	+	6.00	107.41	262.68	364	!	
/[illegible]/2/ina/5B_COMB.seq:US-08-776-971-1[illegible]	+	6.00	107.34	262.68	364	!	
/[illegible]/2/ina/5A_COMB.seq:US-08-203-534-1	+	6.00	104.34	439.18	492	!	S
/[illegible]/2/ina/5A_COMB.seq:US-08-[illegible]	+	6.00	102.28	412.50	574	!	S
/[illegible]/2/ina/5A_COMB.seq:US-08-[illegible]	+	6.00	101.41	413.7	574	!	S
/[illegible]/2/ina/5B_COMB.seq:US-08-[illegible]	+	6.00	101.34	484.15	714	!	
/[illegible]/2/ina/5A_COMB.seq:US-08-471-724-9	+	6.00	101.66	478.94	741	!	S
/[illegible]/2/ina/5B_COMB.seq:US-08-471-969-6	+	6.00	101.66	478.94	741	!	S
/[illegible]/2/ina/5B_COMB.seq:US-08-[illegible]	+	6.00	101.66	478.94	741	!	S
/[illegible]/2/ina/5B_COMB.seq:US-08-470-006A-6	+	6.00	101.66	478.94	741	!	
/[illegible]/2/ina/5A_COMB.seq:US-08-[illegible]	+	6.00	101.66	478.94	741	!	S
/[illegible]/2/ina/5B_COMB.seq:US-[illegible]	+	6.00	100.83	529.72	833	!	
/[illegible]/2/ina/5B_COMB.seq:US-08-[illegible]	+	6.00	100.36	563.47	808	!	
/[illegible]/2/ina/5B_COMB.seq:US-08-[illegible]	+	6.00	99.77	610.2[illegible]	960	!	
/[illegible]/2/ina/5A_COMB.seq:US-08-[illegible]	+	6.00	99.13	637.79	1011	!	
/[illegible]/2/ina/5A_COMB.seq:US-09-[illegible]	+	6.00	99.13	637.79	1011	!	
/[illegible]/2/ina/5A_COMB.seq:US-08-[illegible]	+	6.00	99.05	643.44	1012	!	
/[illegible]/2/ina/5A_COMB.seq:US-08-[illegible]	+	6.00	99.05	643.44	1012	!	
/[illegible]/2/ina/5B_COMB.seq:US-09-[illegible]	+	6.00	99.05	643.44	1012	!	
/[illegible]/2/ina/5A_COMB.seq:US-09-[illegible]	+	6.00	98.92	652.[illegible]	1152	!	
/[illegible]/2/ina/5B_COMB.seq:US-08-[illegible]	+	6.00	98.75	664.84	1151	!	
/[illegible]/2/ina/5B_COMB.seq:US-08-[illegible]	+	6.00	98.10	756.07	1272	!	
/[illegible]/2/ina/5A_COMB.seq:US-09-[illegible]	+	6.00	98.10	756.07	1272	!	
/[illegible]/2/ina/5B_COMB.seq:US-09-[illegible]	+	6.00	98.10	756.07	1272	!	
/[illegible]/2/ina/5A_COMB.seq:US-08-[illegible]	+	6.00	97.73	762.30	1441	!	
/[illegible]/2/ina/5A_COMB.seq:US-08-[illegible]	+	6.00	97.64	802.63	1364	!	
/[illegible]/2/ina/5B_COMB.seq:PCT-US97-[illegible]	+	6.00	97.63	802.63	1364	!	
/[illegible]/2/ina/5A_COMB.seq:US-08-[illegible]	+	6.00	97.57	808.49	1422	!	
/[illegible]/2/ina/5B_COMB.seq:US-08-929-922B-1	+	6.00	97.42	824.32	1409	!	
/[illegible]/2/ina/5A_COMB.seq:US-09-[illegible]	+	6.00	97.42	824.32	1409	!	
/[illegible]/2/ina/5A_COMB.seq:US-08-[illegible]	+	6.00	97.42	824.32	1409	!	
/[illegible]/2/ina/5A_COMB.seq:US-08-640-502-6	+	6.00	97.34	832.72	1426	!	
/[illegible]/2/ina/5B_COMB.seq:US-08-714-991-6	+	6.00	97.34	832.72	1426	!	
/[illegible]/2/ina/5A_COMB.seq:US-08-[illegible]	+	6.00	97.34	832.72	1426	!	
/[illegible]/2/ina/5A_COMB.seq:US-08-[illegible]	+	6.00	97.14	856.43	1473	!	
/[illegible]/2/ina/5B_COMB.seq:US-08-909-965C-13	+	6.00	96.99	871.54	1505	!	
/[illegible]/2/ina/5A_COMB.seq:US-08-[illegible]A-12	+	6.00	96.96	874.08	1510	!	
/[illegible]/2/ina/5A_COMB.seq:US-08-476-500B-1	+	6.00	96.96	874.08	1512	!	
/[illegible]/2/ina/5B_COMB.seq:US-09-[illegible]-5[illegible]-1	+	6.00	96.94	881.41	1525	!	
/[illegible]/2/ina/5A_COMB.seq:US-08-[illegible]-A-6	+	6.00	96.89	885.71	1534	!	
/[illegible]/2/ina/5B_COMB.seq:US-08-714-991-28	+	6.00	96.74	898.99	1561	!	
/[illegible]/2/ina/5B_COMB.seq:US-08-729-979A-2	+	6.00	96.73	900.27	1560	!	
/[illegible]/2/ina/5A_COMB.seq:US-08-168-946-1	+	6.00	96.72	900.91	1565	!	

Sequence [illegible]/2/ina/5B_COMB.seq:US-08-776-971-46

seq_documentation_block:
; Sequence 46, Application US/08776971B
; Patent No. 5228984
; GENERAL INFORMATION:
; APPLICANT: Hinuma, Shuji
; Habata, Yuko
; Kawamata, Yuji
; Hosoya, Masaki
; Fujii, Eyo
; Fukusumi, Shoji
; Kitada, Chieko
; TITLE OF INVENTION: POLYPEPTIDES, THEIR PRODUCTION AND USE
; NUMBER OF SEQUENCES: 140
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08,776,971B
; FILING DATE: 06-Feb-1997
; CLASSIFICATION: Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP95/03821
; FILING DATE: 28-DEC-1995
; APPLICATION NUMBER: JP 7/343371
; FILING DATE: 28-DEC-1995
; APPLICATION NUMBER: JP 8/59419
; FILING DATE: 15-MAR-1996
; APPLICATION NUMBER: JP 8/211805
; FILING DATE: 12 AUG-1996
; APPLICATION NUMBER: JP 8/246573
; FILING DATE: 18-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Conlin, David G.
; REGISTRATION NUMBER: 27,026
; REFERENCE/DOCKET NUMBER: 47176
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617 523-6400
; TELEFAX: 617-523-6440
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 249 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; DESCRIPTION: SEQ ID NO: 46:
US-08-776-971-46

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```
alignment block:
 US-09-126-945B-2_COPY_75_83 x US-08-776-971-46  ..

 Align seq 1/1  to: US-08-776-971-46  from: 1  to: 249

        4 ProGlyAlaSerSerArg 9
          ||||||||||||||||||
       52 CCAGGGGCTTCCAGCGGA 69

seq_name: /usr/prodata/tmp/58_COMB_seq.US-08-776-971-123

seq_documentation block:
; Sequence 123, Application US/08776971B
; Patent No. 6228984
;   GENERAL INFORMATION:
;        APPLICANT: Hinuma, Shuji
;                   Habata, Yuko
;                   Kawamata, Yuji
;                   Bosoya, Masaki
;                   Fujii, Ryo
;                   Fukusumi, Shoji
;                   Kitada, Chieko
;        TITLE OF INVENTION: POLYPROTEINS, THEIR PRODUCTION AND USE
;        NUMBER OF SEQUENCES: 140
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;             STREET: 130 Water Street
;             CITY: Boston
;             STATE: MA
;             COUNTRY: USA
;             ZIP: 02109
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM compatible
;             OPERATING SYSTEM: DOS
;             SOFTWARE: FastSEQ for Windows Version 2.0
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/776,971B
;             FILING DATE: 06-Feb-1997
;             CLASSIFICATION: -Unknown-
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: PCT/JP96/03821
;             FILING DATE: 28-DEC-1996
;             APPLICATION NUMBER: JP 7/343471
;             FILING DATE: 28-DEC-1995
;             APPLICATION NUMBER: JP 8/59419
;             FILING DATE: 15-MAR-1996
;             APPLICATION NUMBER: JP 8/211805
;             FILING DATE: 12-AUG-1996
;             APPLICATION NUMBER: JP 8/246573
;             FILING DATE: 18-SEP-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Conlin, David G.
;             REGISTRATION NUMBER: 27,026
;             REFERENCE/DOCKET NUMBER: 47176
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 617-523-3400
;             TELEFAX: 617 523 6440
;   INFORMATION FOR SEQ ID NO: 123:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 464 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        FEATURE:
;             NAME/KEY:  Coding Sequence
;             LOCATION:  24...272
;             OTHER INFORMATION:
;        SEQUENCE DESCRIPTION: SEQ ID NO: 123:
US-08-776-971-123


alignment scores:
```

```
            Quality:    6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_75_83 x US-08-776-971-123  ..

 Align seq 1/1  to: US-08-776-971-123  from: 1  to: 464

        1 ProGlyAlaSerSerArg 9
          ||||||||||||||||||
       75 CCAGGGGCTTCCAGCGGA 92

seq_name: /usr/prodata/tmp/58_COMB_seq.US-08-776-971-133

seq_documentation block:
; Sequence 133, Application US/08776971B
; Patent No. 6228984
;   GENERAL INFORMATION:
;        APPLICANT: Hinuma, Shuji
;                   Habata, Yuko
;                   Kawamata, Yuji
;                   Bosoya, Masaki
;                   Fujii, Ryo
;                   Fukusumi, Shoji
;                   Kitada, Chieko
;        TITLE OF INVENTION: POLYPROTEINS, THEIR PRODUCTION AND USE
;        NUMBER OF SEQUENCES: 140
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;             STREET: 130 Water Street
;             CITY: Boston
;             STATE: MA
;             COUNTRY: USA
;             ZIP: 02109
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM compatible
;             OPERATING SYSTEM: DOS
;             SOFTWARE: FastSEQ for Windows Version 2.0
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/776,971B
;             FILING DATE: 06-Feb-1997
;             CLASSIFICATION: -Unknown-
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: PCT/JP96/03821
;             FILING DATE: 28-DEC-1996
;             APPLICATION NUMBER: JP 7/343471
;             FILING DATE: 28-DEC-1995
;             APPLICATION NUMBER: JP 8/59419
;             FILING DATE: 15-MAR-1996
;             APPLICATION NUMBER: JP 8/211805
;             FILING DATE: 12-AUG-1996
;             APPLICATION NUMBER: JP 8/246573
;             FILING DATE: 18-SEP-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Conlin, David G.
;             REGISTRATION NUMBER: 27,026
;             REFERENCE/DOCKET NUMBER: 47176
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 617-523-3400
;             TELEFAX: 617 523 6440
;   INFORMATION FOR SEQ ID NO: 133:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 464 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 133:
US-08-776-971-133


alignment scores:
```

```
            Quality:     6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_75_83 x US-08-776-971-133

 Align seq 1/1  to: US-08-776-971-133  from: 3  to: 94

       4 ProGlyAlaSerSerArg 9
         |||||||||||||||||
      75 CCAGGGGCTTCAAGCCGA 92

seq_name: /cmc2_2/p4_fra/2/ina/5A_COMB.seq:US-08-776-971-133 1

seq_documentation_block:
; Sequence 1, Application US/08/203534
; Patent No. 5518884
;  GENERAL INFORMATION:
;    APPLICANT:  Spears, Patricia A.
;    APPLICANT:  Shank, Daryl D.
;    TITLE OF INVENTION:  NUCLEIC ACID SEQUENCES SPECIFIC FOR
;    TITLE OF INVENTION:  MYCOBACTERIUM KANSASII
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Richard J. Patrick, Becton Dickinson and
;      ADDRESSEE:  Company
;      STREET:  1 Becton Drive
;      CITY:  Franklin Lakes
;      STATE:  NJ
;      COUNTRY:  US
;      ZIP:  07417
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/203,534
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fuqit, Donna R.
;      REGISTRATION NUMBER:  32,135
;      REFERENCE/DOCKET NUMBER:  P-2858
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  493 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Mycobacterium kansasii
US-08-203-534-1


alignment_scores:
            Quality:     6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_75_83 x US-08-203-534-1/rev  ..

 Align seq 1/1  to reverse of: US-08-203-534-1  from: 1  to: 493

       2 LysAlaProGlyAlaSer 7
         ||||||||||||||||||
     105 AAGGCTCCGGGGGCAAGT 88

seq_name: /cmc2_2/p4_fra/2/ina/5A_COMB.seq:US-08-203-534-1 1

seq_documentation_block:
; Sequence 1, Application US/09069811
; Patent No. 6068974
;  GENERAL INFORMATION:
;    APPLICANT:  Klamm, Richard C.
;    TITLE OF INVENTION:  A SPECIFIC, HIGHLY SENSITIVE, NESTED PCR
;    TITLE OF INVENTION:  DETECTION SCHEME FOR THE PSEUDORABIES VIRUS
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  W. Murray Spruill (Alston & Bird, LLP)
;      STREET:  3605 Glenwood Ave, Suite 410
;      CITY:  Raleigh
;      STATE:  NC
;      COUNTRY:  US
;      ZIP:  27622
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/069,811
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spruill, W. Murray
;      REGISTRATION NUMBER:  32,943
;      REFERENCE/DOCKET NUMBER:  5626-16
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919 420 2202
;      TELEFAX:  919 881 3175
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  674 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      [illegible] N  /desc=  "Region of the pseudorabies
;      DESCRIPTION:  virus gII gene"
;    ORIGINAL SOURCE:
;      ORGANISM:  Pseudorabies virus
US-09-069-811-1


alignment_scores:
            Quality:     6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_75_83 x US-09-069-811-1  ..

 Align seq 1/1  to: US-09-069-811-1  from: 1  to: 674

       4 ProGlyAlaSerSerArg 9
         |||||||||||||||||
     485 CCGGGGCTTCAAGCAGG 502

seq_name: /cmc2_2/p4_fra/2/ina/5A_COMB.seq:US-08-813-942-1 1

seq_documentation_block:
; Sequence 1, Application US/08613942
; Patent No. 5866791
;  GENERAL INFORMATION:
;    APPLICANT:  BELL, KAREN A.
;    TITLE OF INVENTION:  MODIFICATION OF LIGNIN SYNTHESIS IN
;    TITLE OF INVENTION:  PLANTS
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  [illegible] Ave., N.W.
;      CITY:  Washington
```

```
;       STATE:  D.C.
;       COUNTRY:  U.S.A.
;       ZIP:  20005-3918
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/613,941
;       FILING DATE:  14-MAR-1996
;       CLASSIFICATION:  800
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  GB 9521048.3
;       FILING DATE:  25-OCT-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  KOKOLIS, PAUL N.
;       REGISTRATION NUMBER:  16,773
;       REFERENCE/DOCKET NUMBER:  22234/SEE 45024
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 861-3000
;       TELEFAX:  (202) 822-0944
;       TELEX:  6714627 CUSH
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  714 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  both
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-613-942-1


alignment scores:
            Quality:     6.00              Length:       6
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_75_83 x US-08-613-942-1/rev   ..

 Align Seq 1/1  to reverse of: US-08-613-942-1  from: 1  to: 714

        8 AlaProGlyAlaSerSer 8
          |||||||||||||||||
      277 GCTCCCGGCGCCTCCAGC 260

Seq name: /equiv/ftptmpdata/zzzrna/6B_COMB.seq:US-08-998-416-916

Seq documentation block:
; Sequence 916, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;     APPLICANT:  Philippsen, Peter
;     APPLICANT:  Pohlmann, Rainer
;     APPLICANT:  Steiner, Sabine
;     APPLICANT:  Mohr, Christine
;     APPLICANT:  Wendland, Jurgen
;     APPLICANT:  Knechtle, Philipp
;     APPLICANT:  Robischung, Corinne
;     TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;     TITLE OF INVENTION:  AND USES THEREOF
;     NUMBER OF SEQUENCES:  1152
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  No. 6239264arris Corporation
;       STREET:  3054 Cornwallis Road
;       CITY:  Research Triangle Park
;       STATE:  No. 6239264th Carolina
;       COUNTRY:  USA
;       ZIP:  27709
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/998,416
;       FILING DATE:  24-DEC-1997
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  CH 0016/97
;       FILING DATE:  31-DEC-1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Mehta, J. Timothy
;       REGISTRATION NUMBER:  38,241
;       REFERENCE/DOCKET NUMBER:  PF/[illegible]
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  919-541-8587
;       TELEFAX:  919-541-8689
;   INFORMATION FOR SEQ ID NO:  916:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  714 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  simple
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     ORIGINAL SOURCE:
;       ORGANISM:  PAG1570DP
US-08-998-416-916


alignment scores:
            Quality:     6.00              Length:       6
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_75_83 x US-08-998-416-916/rev   ..

 Align Seq 1/1  to reverse of: US-08-998-416-916  from: 1  to: 714

        4 ProGlyAlaSerSerArg 9
          |||||||||||||||||
       92 CCAGGTGCGAGTTCACGA 75

Seq name: /[illegible]/ftptmpdata/zzzrna/6B_COMB.seq:US-08-471-724-9

Seq documentation block:
; Sequence 9, Application US/08471724
; Patent No. 5800980
;  GENERAL INFORMATION:
;     APPLICANT:  Herve PERRON
;     APPLICANT:  Francois MALLET
;     APPLICANT:  Bernard MANDRAND
;     APPLICANT:  Frederic BEDIN
;     APPLICANT:  Frederic BESEME
;     TITLE OF INVENTION:  MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR
;     TITLE OF INVENTION:  INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND BIO
;     TITLE OF INVENTION:  CONSTITUENTS THEREOF (AS AMENDED)
;     NUMBER OF SEQUENCES:  48
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Olitt & Berridge
;       STREET:  700 South Washington Street, Suite 600
;       CITY:  Alexandria
;       STATE:  Virginia
;       COUNTRY:  U.S.A.
;       ZIP:  22314
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/471,724
;       FILING DATE:  June 6, 1995
;     ATTORNEY/AGENT INFORMATION:
```

```
;       NAME:  Berridge, William P.
;       REGISTRATION NUMBER:  30,024
;       REFERENCE/DOCKET NUMBER:  WPB 36055C
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  703-836-6400
;       TELEFAX:  703-836-2787
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  741 bases
;       TYPE:  nucleotide
;       STRANDEDNESS:  single-stranded
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-471-724-9


alignment_scores:
             Quality:    6.00              Length:       6
               Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_75_83 x US-08-471-724-9    ..

 Align seq 1/1  to: US-08-471-724-9  from: 1  to: 741

       2 LysAlaProGlyAlaSer 7
         ||||||||||||||||||
     113 AAGGCACCAGGGGCCTCA 130

seq_name: /cm2_fp1_dir2/ind758_2_COPY_seq_US-09-471-969-9

seq_documentation_block:
; Sequence 9, Application US/08471969
; Patent No. 5871745
;  GENERAL INFORMATION:
;     APPLICANT:  Herve PERRON
;     APPLICANT:  Francois MALLET
;     APPLICANT:  Bernard MANDRAND
;     APPLICANT:  Frederic BEDIN
;     APPLICANT:  Frederic BESEME
;     TITLE OF INVENTION:  MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR
;     TITLE OF INVENTION:  INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND BIOPOL
;     TITLE OF INVENTION:  CONSTITUENTS THEREOF (AS AMENDED)
;     NUMBER OF SEQUENCES:  38
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Olitt & Berridge
;       STREET:  700 South Washington Street, Suite 400
;       CITY:  Alexandria
;       STATE:  Virginia
;       COUNTRY:  U.S.A.
;       ZIP:  22314
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/471,969
;       FILING DATE:  June 6, 1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Berridge, William P.
;       REGISTRATION NUMBER:  30,024
;       REFERENCE/DOCKET NUMBER:  WPB 36055A
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  703-836-6400
;       TELEFAX:  703-836-2787
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  741 bases
;       TYPE:  nucleotide
;       STRANDEDNESS:  single-stranded
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-471-969-9


alignment_scores:
             Quality:    6.00              Length:       6
               Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_75_83 x US-08-471-969-9    ..

 Align seq 1/1  to: US-08-471-969-9  from: 1  to: 741

       2 LysAlaProGlyAlaSer 7
         ||||||||||||||||||
     113 AAGGCACCAGGGGCCTCA 130

seq_name: /cm2_fp1_dir2/ind758_2_COPY_seq_US-09-471-969-9

seq_documentation_block:
; Sequence 9, Application US/08384137
; Patent No. 5871996
;  GENERAL INFORMATION:
;     APPLICANT:  Herve PERRON
;     APPLICANT:  Francois MALLET
;     APPLICANT:  Bernard MANDRAND
;     APPLICANT:  Frederic BEDIN
;     APPLICANT:  Frederic BESEME
;     TITLE OF INVENTION:  MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR
;     TITLE OF INVENTION:  INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND BIO
;     TITLE OF INVENTION:  CONSTITUENTS THEREOF (AS AMENDED)
;     NUMBER OF SEQUENCES:  38
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Olitt & Berridge
;       STREET:  700 South Washington Street, Suite 400
;       CITY:  Alexandria
;       STATE:  Virginia
;       COUNTRY:  U.S.A.
;       ZIP:  22314
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/384,137
;       FILING DATE:  February 6, 1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Berridge, William P.
;       REGISTRATION NUMBER:  30,024
;       REFERENCE/DOCKET NUMBER:  WPB 36055
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  703 836-6400
;       TELEFAX:  703-836-2787
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  741 bases
;       TYPE:  nucleotide
;       STRANDEDNESS:  single stranded
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-384-137-9


alignment_scores:
             Quality:    6.00              Length:       6
               Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_75_83 x US-08-384-137-9    ..
```

```
 Align seq 1/1  to: US-08-691-564C-9  from: 1  to: 741

        2 LysAlaProGlyAlaSer 7
          ||||||||||||||||||
      113 AAGGCACCAGGGGCCTCA 130

seq name: /cgn2_6/ptodata/2/iia/%8_COMB.seq:US-08-691-564C-9 ; ver 1.0 ; A

seq documentation block:
; Sequence 9, Application US/08470006A
; Patent No. 5962217
;  GENERAL INFORMATION:
;     APPLICANT:  Herve PERRON
;     APPLICANT:  Francois MALLET
;     APPLICANT:  Bernard MANDRAND
;     APPLICANT:  Frederic BEDIN
;     APPLICANT:  Frederic BESEME
;     TITLE OF INVENTION:  MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR
;     TITLE OF INVENTION:  INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND
;     TITLE OF INVENTION:  BIOPOLYMER CONSTITUENTS THEREOF (AS AMENDED)
;     NUMBER OF SEQUENCES:  48
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Olitt & Berridge
;       STREET:  700 South Washington Street, Suite 300
;       CITY:  Alexandria
;       STATE:  Virginia
;       COUNTRY:  U.S.A.
;       ZIP:  22314
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/470,006A
;       FILING DATE:  June 6, 1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Berridge, William P.
;       REGISTRATION NUMBER:  30,024
;       REFERENCE/DOCKET NUMBER:  WPB 36055B
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  703 836 6400
;       TELEFAX:  703 836 2787
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  741 bases
;       TYPE:  nucleotide
;       STRANDEDNESS:  single stranded
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-470-006A-9


alignment_scores:
             Quality:    6.00              Length:       6
               Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_75_83 x US-08-470-006A-9  ..

 Align seq 1/1  to: US-08-470-006A-9  from: 1  to: 741

        2 LysAlaProGlyAlaSer 7
          ||||||||||||||||||
      113 AAGGCACCAGGGGCCTCA 130

seq name: /cgn2_6/ptodata/2/iia/6A_COMB.seq:US-08-691-564C-9

seq documentation block:
; Sequence 9, Application US/08691564C
; Patent No. 6001987
;  GENERAL INFORMATION:
;     APPLICANT:  Herve PERRON
;     APPLICANT:  Frederic BESEME
;     APPLICANT:  Frederic BEDIN
;     APPLICANT:  Glaucia PAKANHOS BACCALA
;     APPLICANT:  Florence KOMURIAN-PRADEL
;     APPLICANT:  Colette JOLIVET
;     APPLICANT:  Bernard MANDRAND
;     TITLE OF INVENTION:  VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;     TITLE OF INVENTION:  ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
;     TITLE OF INVENTION:  THERAPEUTIC PURPOSES
;     NUMBER OF SEQUENCES:  92
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Olitt & Berridge
;       STREET:  700 South Washington Street, Suite 300
;       CITY:  Alexandria
;       STATE:  Virginia
;       COUNTRY:  U.S.A.
;       ZIP:  22314
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/691,563C
;       FILING DATE:  02-AUG-1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Berridge, William P.
;       REGISTRATION NUMBER:  30,024
;       REFERENCE/DOCKET NUMBER:  WPB 48588
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  703 836 6400
;       TELEFAX:  703 836 2787
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  741 base pairs
;       TYPE:  nucleotide
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-691-563C-9


alignment_scores:
             Quality:    6.00              Length:       6
               Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_75_83 x US-08-691-563C-9  ..

 Align seq 1/1  to: US-08-691-563C-9  from: 1  to: 741

        2 LysAlaProGlyAlaSer 7
          ||||||||||||||||||
      113 AAGGCACCAGGGGCCTCA 130

seq name: /cgn_6/ptodata/2/iia/6A_COMB.seq:US-09-200-990-9

seq documentation block:
; Sequence 9, Application US/09200990
; Patent No. 6184025
;  GENERAL INFORMATION:
;     APPLICANT:  Herve PERRON
;     APPLICANT:  Francois MALLET
;     APPLICANT:  Bernard MANDRAND
;     APPLICANT:  Frederic BEDIN
;     APPLICANT:  Frederic BESEME
;     TITLE OF INVENTION:  MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR
;     TITLE OF INVENTION:  INFECTIVE AGENT ASSOCIATED WITH MULITPLE SCLEROSIS, AND BIO
;     TITLE OF INVENTION:  CONSTITUENTS THEREOF (AS AMENDED)
;     NUMBER OF SEQUENCES:  48
;     CORRESPONDENCE ADDRESS:
```

```
;       ADDRESSEE:  Olitt & Berridge
;       STREET:  700 South Washington Street  Suite 500
;       CITY:  Alexandria
;       STATE:  Virginia
;       COUNTRY:  U.S.A.
;       ZIP:  22314
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/200,990
;       FILING DATE:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/471,969
;       FILING DATE:  June 6, 1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Berridge, William P.
;       REGISTRATION NUMBER:  30,024
;       REFERENCE/DOCKET NUMBER:  WPB 36055A
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  703-836-6400
;       TELEFAX:  703-836-2787
;   INFORMATION FOR SEQ ID NO:  99:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  741 bases
;       TYPE:  nucleotide
;       STRANDEDNESS:  single-stranded
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-09-200-990-9


alignment_scores:
            Quality:     6.00              Length:       6
              Ratio:  1.000                 Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_75_83 x US-09-200-990-9  ..

 Align seq 1/1  to: US-09-200-990-9  from: 1  to: 741

        2 LysAlaProGlyAlaSer 7
          ||||||||||||||||||
      113 AAGGCACCAGGGGCCTCA 130

Seq_name: /mnt/ftptemp/zhang/5B_COMB.seq:US-08-837-029-1

Seq_documentation_block:
; Sequence 1, Application US/08837029
; Patent No. 5945303
;  GENERAL INFORMATION:
;     APPLICANT:  Wei et al.
;     TITLE OF INVENTION:  Human Hematopoietic - Specific Protein
;     NUMBER OF SEQUENCES:  8
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Human Genome Sciences, Inc.
;       STREET:  9410 Key West Avenue
;       CITY:  Rockville
;       STATE:  MD
;       COUNTRY:  US
;       ZIP:  20850
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/837,029
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/966,316
;       FILING DATE:  11-APR-1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Brookes, A. Anders
;       REGISTRATION NUMBER:  36,373
;       REFERENCE/DOCKET NUMBER:  PF268
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  301-309-8504
;       TELEFAX:  301-309-8439
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  833 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  42..608
;     FEATURE:
;       NAME/KEY:  sig_peptide
;       LOCATION:  42..107
;     FEATURE:
;       NAME/KEY:  mat_peptide
;       LOCATION:  108..608
US-08-837-029-1


alignment_scores:
            Quality:     6.00              Length:       6
              Ratio:  1.000                 Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_75_83 x US-08-837-029-1_rev  ..

 Align seq 1/1  to reverse of: US-08-837-029-1  from: 1  to: 833

        3 AlaProGlyAlaSerSer 8
          ||||||||||||||||||
       40 GCCCCAGGAGCCAGTTCA 23

Seq_name: /mnt/ftptemp/zhang/5B_COMB.seq:US-08-966-316-4

Seq_documentation_block:
; Sequence 4, Application US/08966316
; Patent No. 5932445
;  GENERAL INFORMATION:
;     APPLICANT:  Lal, Preeti
;     APPLICANT:  Au-Young, Janice
;     APPLICANT:  Reddy, Roopa
;     APPLICANT:  Murry, Lynn E.
;     APPLICANT:  Mathur, Preeta
;     TITLE OF INVENTION:  SIGNAL PEPTIDE  CONTAINING PROTEINS
;     NUMBER OF SEQUENCES:  18
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;       STREET:  3174 Porter Drive
;       CITY:  Palo Alto
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  94304
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/966,316
;       FILING DATE:  Herewith
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
```

```
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Billings, Lucy J.
;       REGISTRATION NUMBER:  36,749
;       REFERENCE/DOCKET NUMBER:  PF 0424 US
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  650 855 0555
;       TELEFAX:  650 845-4166
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  898 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     IMMEDIATE SOURCE:
;       LIBRARY:  BLADDUI06
;       CLONE:  IB99182
US 08 966 416 4


alignment_scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000               Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US 09 126 945B 2 COPY_75_83 x US 08 966 416 4    ..

 Align seq 1/1  to: US-08-966-416-4  from: 1  to: 898

        3 AlaProGlyAlaSerSer 8
          ||||||||||||||||||
      866 GCACCTGGCCCCTCGAGC 883
```

Run on: US-09-126-945B-2_COPY_75_83.* : N_Geneseq_0601:* *

Date: Sep 14, 2001 9:48 PM

About: Results were produced by the GenCore software version 4.5
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL frame_p2n.model -DEV xlp
-Q "/cgn2_1/tmp/US09_126_945B_2_COPY_75_83.rng" -DB N_Geneseq_0601 -QFMT=fastap -SUFFIX=olip2n.rng -GAPOP 4.500
-GAPEXT 0.050 -MINMATCH=0.100 -LOOPCL 0.000 -LOOPEXT 0.000
-QGAPOP 4.500 -QGAPEXT 0.050 -XGAPOP=60.000 -XGAPEXT 60.000
-FGAPOP 6.000 -FGAPEXT 7.000 -YGAPOP=60.000 -YGAPEXT 60.000
-DELOP 6.000 -DELEXT 7.000 -START 1 -MATRIX oligo
-TRANS human40.cdi -LIST=45 -DOCALIGN 200 -THR_SCORE quality
-THR_MIN 1 -ALIGN=15 -MODE=LOCAL -OUTFMT pir -NORM cxf -MINLEN 9
-MAXLEN 2000000000 -USER=US09126945_COPY_1_608 -NOPH 6 -LOOP 5
-LONGLOG -NO_XLPXY -WAIT -THREADS 1

Search information block:
Query: US-09-126-945B-2_COPY_75_83
Query length: 9
Database: N_Geneseq_0601:*
Database sequences: 740101
Database length: 413950809
Search time (sec): 680.150000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT 60.000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	%	Documentation			
/SIDS1/p/data/geneseq/geneseqpo/NA_60	DAT	AAF_2828	+	6.00	154....	...	...	...	...	...
/SIDS1/p/data/geneseq/geneseqpo/NA2000	DAT	AAF21801	+	6.00	152.87	1.00	1084			
/SIDS1/p/data/geneseq/geneseqpo/NA2002	DAT	AAF83261	+	6.00	152.87	1.00	1084			
/SIDS1/p/data/geneseq/geneseqpo/NA1994	DAT	AAF66015	+	6.00	150.92	1.00	1070			
/SIDS1/p/data/geneseq/geneseqpo/NA2002	DAT	AAF83227	+	6.00	148.17	2.78	1057			
/SIDS1/p/data/geneseq/geneseqpo/NA1999	DAT	AAF23153	+	6.00	141.27	5.56	1004			
/SIDS1/p/data/geneseq/geneseqpo/NA1993	DAT	AAF59397	+	7.00	139.82	29.27	282			
/SIDS1/p/data/geneseq/geneseqpo/NA2001	DAT	AAF11805	+	7.00	132.25	27.35	265			
/SIDS1/p/data/geneseq/geneseqpo/NA2002	DAT	AAF46674	+	7.00	119.89	101.52	1174			
/SIDS1/p/data/geneseq/geneseqpo/NA1993	DAT	AAF17726	+	7.00	119.22	113.94	1292			
/SIDS1/p/data/geneseq/geneseqpo/NA2000	DAT	AAF59248	+	7.00	116.23	167.11	1984			
/SIDS1/p/data/geneseq/geneseqpo/NA2001	DAT	AAF25173	+	7.00	116.23	167.26	1986			
/SIDS1/p/data/geneseq/geneseqpo/NA2000	DAT	AAF15606	+	7.00	114.08	250.22	2719			
/SIDS1/p/data/geneseq/geneseqpo/NA2001	DAT	AAF54962	+	7.00	110.83	356.38	4632			
/SIDS1/p/data/geneseq/geneseqpo/NA2000	DAT	AAA59146	+	7.00	100.48	1.60e+03	19053			
/SIDS1/p/data/geneseq/geneseqpo/NA1999	DAT	AAZ23903	+	7.00	93.77	3.00e+03	49999			
/SIDS1/p/data/geneseq/geneseqpo/NA2000	DAT	AAZ84751	+	7.00	86.92	7.20e+03	133716			
/SIDS1/p/data/geneseq/geneseqpo/NA1997	DAT	AAI78888	+	6.00	129.66	22.96	26			
/SIDS1/p/data/geneseq/geneseqpo/NA1990	DAT	AAI11017	+	6.00	129.76	43.35	51			
/SIDS1/p/data/geneseq/geneseqpo/NA_30	DAT	AAA_25	+	6.00	126.83	92.65	90			
/SIDS1/p/data/geneseq/geneseqpo/NA1984	DAT	AAB13...	+	6.00	119.91	218.22	1.5			
/SIDS1/p/data/geneseq/geneseqpo/NA2000	DAT	AAZ21471	+	6.00	114.77	201.52	191			
/SIDS1/p/data/geneseq/geneseqpo/NA1994	DAT	AAZ55742	+	6.00	113.56	207.17	197			
/SIDS1/p/data/geneseq/geneseqpo/NA2000	DAT	AAF14271	+	6.00	112.95	223.94	212			
/SIDS1/p/data/geneseq/geneseqpo/NA_60	DAT	AAA_224	+	6.00	111.65	231.41	220			
/SIDS1/p/data/geneseq/geneseqpo/NA_07	DAT	AAA15729	+	6.00	111.63	233.27	225			
/SIDS1/p/data/geneseq/geneseqpo/NA_60	DAT	AAA_134	+	6.00	111.63	233.27	225			
/SIDS1/p/data/geneseq/geneseqpo/NA1997	DAT	AAV12...	+	6.00	111.63	255.37	244			
/SIDS1/p/data/geneseq/geneseqpo/NA1994	DAT	AAA15525	+	6.00	111.63	255.37	244			
/SIDS1/p/data/geneseq/geneseqpo/NA1999	DAT	AAV91728	+	6.00	111.63	255.37	244			
/SIDS1/p/data/geneseq/geneseqpo/NA2000	DAT	AAA71438	+	6.00	111.63	255.37	244			
/SIDS1/p/data/geneseq/geneseqpo/NA2003	DAT	AAF12133	+	6.00	111.60	266.32	254			
/SIDS1/p/data/geneseq/geneseqpo/NA1999	DAT	AAV90453	+	6.00	111.46	274.51	262			
/SIDS1/p/data/geneseq/geneseqpo/NA2000	DAT	AAZ27462	+	6.00	111.46	274.51	262			
/SIDS1/p/data/geneseq/geneseqpo/NA2001	DAT	AAZ09784	+	6.00	111.44	275.12	263			
/SIDS1/p/data/geneseq/geneseqpo/NA_03	DAT	AA_2...	+	6.00	111.43	275.23	262			
/SIDS1/p/data/geneseq/geneseqpo/NA_08	DAT	AAZ12...	+	6.00	111.28	278.11	271			
/SIDS1/p/data/geneseq/geneseqpo/NA2000	DAT	AAA01377	+	6.00	111.63	201.60	200			
/SIDS1/p/data/geneseq/geneseqpo/NA1990	DAT	AAA01392	+	6.00	111.61	302.49	301			
/SIDS1/p/data/geneseq/geneseqpo/NA2000	DAT	AAV29740	+	6.00	111.58	311.45	311			
/SIDS1/p/data/geneseq/geneseqpo/NA1993	DAT	AAZ01745	+	6.00	111.27	315.92	316			
/SIDS1/p/data/geneseq/geneseqpo/NA2000	DAT	AAC02824	+	6.00	111.27	326.72	327			
/SIDS1/p/data/geneseq/geneseqpo/NA1995	DAT	AAF23924	+	6.00	110.85	335.49	338			
/SIDS1/p/data/geneseq/geneseqpo/NA2000	DAT	AAV86229	+	6.00	110.46	350.52	356			
/SIDS1/p/data/geneseq/geneseqpo/NA1995	DAT	AAF21892	+	6.00	110.12	362.16	367			

seq_id: /SIDS1/p/data/geneseq/geneseqpo/NA2000 DAT:AAF21828

seq_id #full#:
ID AAF21828 standard; DNA; 1087 BP
XX
AC AAF21828;
XX
DT 27-MAR-2001 (first entry)
XX
DE Human breast and ovarian cancer associated antigen gene SEQ ID 215.
XX
KW Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
KW nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
KW antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
KW Addison's disease; allergy; autoimmune haemolytic anaemia;
KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW cardiovascular disorder; wound healing; neurological disease; ds.
XX
OS Homo sapiens.
XX
PN WO200055173-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05881.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Kasaian M; Polokoff MA;
XX
DR WPI; 2000-610715/58.
DR P-PSDB; AAB58925.
XX
PT New human breast and ovarian cancer associated gene sequences and the
PT polypeptides encoded by these genes, useful in the prevention,
PT treatment and diagnosis of cancer, immune disorders, cardiovascular
PT disorders and neurological diseases
XX
PS Claim 1; Page 642-643; 1299pp; English.
XX
CC Sequences AAF21811 - AAF22041 represent DNA sequences encoding human
CC proteins, AAB58711 - AAB59128. The DNA and protein sequences are
CC associated with breast and ovarian cancer. Included in the invention are
CC sequences AAF21... - AAF22041 and AAB59123 which are used in the
CC isolation and characterisation of the DNA and protein sequences of the
CC invention. The breast and ovarian cancer associated DNA, protein, agonist
CC and antagonist sequences exhibit cytostatic; immunosuppressive;
CC nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
CC antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
CC antibacterial, antifungal, antiparasitic and cardiant activity. The
CC polypeptide and protein sequences are used in the diagnosis, treatment,
CC particularly of breast and ovarian cancer. The nucleic acid sequences,
CC proteins, agonists and antagonists may also be used in the diagnosis,
CC prevention and treatment of immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis, cardiovascular disorders such as
CC cardiac ischaemia, wound healing, neurological diseases such as
CC cerebral ischaemia and epilepsy, and infectious diseases.
XX
SQ Sequence 1087 BP; 215 A; 347 C; 339 G; 182 T; 4 other;

```
alignment scores:
            Quality:     9.00              Length:       9
              Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2 COPY 75-83 x AAF21828    ..

 Align Seq 1/1   to: AAF21828   from: 1   to: 1097

        1 AlaLysAlaProGlyAlaSerSerArg 9
          ||||||||||||||||||||||||||||
       99 GCCAAGGCCCCTGGGGCCAGCAGTCGG 125

seq name: /sli6s1/seqdata4/geneseq2genoseqp2/NA2000.DAT:AAZ50691

seq_documentation block:
ID   AAZ50691 standard; cDNA; 1894 BP.
XX
AC   AAZ50691;
XX
DT   23-MAY-2000  (first entry)
XX
DE   Human prostate derived Ets factor cDNA.
XX
KW   Prostate Derived Ets factor; PDEF; human; chromosome 6p21.3; cancer;
KW   loss of heterozygosity; chromosomal translocation; linkage analysis;
KW   cytostatic; cardiant; immunosuppressive; cerebroprotective; tunicicide;
KW   antibacterial; vulnerary; neuroprotective; antiparkinsonian; nootropic;
KW   anabolic; antiinflammatory; anorectic; hybridisation probe; forensic;
KW   tumour marker; diagnosis, treatment, prostate cancer, blood coagulation,
KW   autoimmune disorder; haematopoietic; immune/nervous system; stroke;
KW   neoplasm; microbial infection; tissue regeneration; heart attack;
KW   scarring; food additive; preservative; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             416..1423
FT                   /*tag= a
FT                   /product= "Human prostate derived Ets factor"
FT                   /note= "Expressed in prostate epithelium"
XX
PN   WO200006589-A1.
XX
PD   10-FEB-2000.
XX
PF   02-AUG-1999;   99WO-US17470.
XX
PR   31-JUL-1998;   98US-012694P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT
XX
PI   Libermann TA; Oettgen JP; Kunsch CA; Endress GA; Rosen CA;
XX
DR   WPI; 2000-195255/17.
DR   P-PSDB; AAY45003.
XX
PT   Novel prostate derived polypeptide, polynucleotide useful for
PT   diagnosis, prevention and treatment of prostate cancer, autoimmune
PT   disorders, microbial infections and also as food additive or
PT   preservative
XX
PS   Claim 4; Fig 1; 142pp; English.
XX
CC   The present cDNA sequence encodes a human prostate derived Ets factor
CC   (PDEF), isolated from human prostate carcinoma cDNA library
CC   (APCC No.203072). PDEF gene expression is highest in tissues involved in
CC   androgen sensitivity, with enriched expression in prostate epithelium
CC   cells. PDEF gene is mapped to the human chromosome 6p21.3
CC   region that is associated with loss of heterozygosity and chromosomal
CC   translocations in various human cancers. PDEF has cytostatic, cardiant,
CC   immunosuppressive, cerebroprotective, tunicicide, antibacterial,
CC   vulnerary, neuroprotective, antiparkinsonian, nootropic, anabolic,
CC   antiinflammatory and anorectic activity. PDEF polynucleotides are useful
CC   in linkage analysis as markers, as hybridisation probes for differential
CC   identification of the tissues or cell types and as polymorphic markers
CC   for forensic purposes. PDEF is useful as prostate specific tumour marker
CC   for the diagnosis and treatment of prostate cancer. PDEF sequences are
CC   useful for treating autoimmune disorders, haematopoietic, blood
CC   coagulation, immune and nervous system disorders, hyperproliferative
CC   disorders like, neoplasms and microbial infections, heart attacks,
CC   stroke, scarring and for tissue regeneration. They are also useful as
CC   food additives or preservatives.
XX
SQ   Sequence 1894 BP; 368 A; 654 C; 571 G; 302 T; 0 other;


alignment scores:
            Quality:     9.00              Length:       9
              Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2 COPY 75-83 x AAZ50691    ..

 Align Seq 1/1   to: AAZ50691   from: 1   to: 1894

        1 AlaLysAlaProGlyAlaSerSerArg 9
          ||||||||||||||||||||||||||||
      638 GCCAAGGCCCCTGGGGCCAGCAGTCGG 664

seq name: /sli6s1/seqdata4/geneseq2genoseqp2/NA2001.DAT:AAC83261

seq_documentation block:
ID   AAC83261 standard; DNA; 1894 BP.
XX
AC   AAC83261;
XX
DT   16-MAR-2001  (first entry)
XX
DE   Gland specific Ets transcription factor (GSEF) nucleotide sequence.
XX
KW   Transcription factor; Gland specific Ets transcription factor; GSEF;
KW   metastatic potential; cancer; tumour; metastasis; breast; prostate;
KW   leukaemia; lymphoma; sarcoma; melanoma; chromosome 6p21.1 6p21.3; ds.
XX
OS   Homo sapiens.
XX
PN   WO200073779-A1.
XX
PD   23-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US13173.
XX
PR   14-MAY-1999;   99US-0134112.
XX
PA   (CHIR-) CHIRON CORP.
XX
PI   Kaufmann J; Xin H; Harrowe G;
XX
DR   WPI; 2001-041197/05.
DR   P-PSDB; AAB49028.
XX
PT   Detecting metastatic and potential metastatic cancerous cells, useful
PT   for diagnosing, prognosing, grading and staging of cancers by detecting
PT   gland specific Ets transcription factor gene product in a biological
PT   sample from a cell
XX
PS   Disclosure; Page 81-83; 95pp; English.
XX
CC   This invention relates to a method for the detection and determination of
CC   the metastatic potential of a cell. The method comprises detecting a
CC   gland specific Ets transcription factor (GSEF) gene product in a test
```

```
CC   sample. Detection of a GSEF gene product in the test sample in amount
CC   lower than that in a normal cell, is indicative of a cell with high
CC   metastatic potential. The method is useful for determining the metastatic
CC   potential of a cell, for the diagnosis and prognosis of cancer as well as
CC   grading and staging of cancers by detecting GSEF expression in a
CC   biological test sample. The method may also be used to monitor patients
CC   having a predisposition to develop a particular cancer. GSEF polypeptides
CC   are useful for producing antibodies, in cancer diagnosis, prognosis,
CC   grading, staging and management of breast and prostate tumours, and in
CC   detecting polymorphisms in the sequence. GSEF genes and proteins are also
CC   useful in gene therapy. GSEF gene product expression levels can be used
CC   in conjunction with any tissue in which an alteration in GSEF gene
CC   product expression levels is associated with development of a
CC   cancer associated phenotype. Cancers, which can be monitored include
CC   cancers of the prostate, cervix, lung and colon, melanomas, colorectal
CC   adenocarcinomas, Wilm's tumour, retinoblastoma, sarcomas, myosarcomas,
CC   lung carcinomas, leukaemia, and lymphomas. The GSEF gene is located on
CC   human chromosome 6, specifically at 6p21.1 6p21.3. The present sequence
CC   represents the DNA encoding GSEF.
XX
SQ   Sequence 1894 BP; 368 A; 654 C; 571 G; 302 T; 0 other;


alignment_scores:
              Quality:    9.00              Length:       9
                Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_75_83 x AAC83261   ..

 Align seq 1/1  to: AAC83261  from: 1  to: 1894

        1 AlaLysAlaProGlyAlaSerSerArg 9
          ||||||||||||||||||||||||||||
      638 GCCAAGGCCCCTGGGGCCAGCAGTCGG 664

seq_name: ; ID: ; locality/genus/species/genotype; NA: 64 ; AC: AAC83261

seq_documentation_block:
ID   AAC83261 standard; DNA; 1894 BP.
XX
AC   AAC83261;
XX
DT   21-NOV-1994  (first entry)
XX
DE   Wnt clone contg. the HR2 gene.
XX
KW   Haematopoietic regulatory protein; differentiation; HPC;
KW   haematopoietic progenitor cells; development; leukaemia; lymphoma;
KW   anaemia; amplification; PCR; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             342..1501
FT                   /*tag= a
XX
PN   WO9411510-A.
XX
PD   26-MAY-1994.
XX
PF   15-NOV-1993;   93WO-US10728.
XX
PR   16-NOV-1992;   92US-0976789.
XX
PA   (SCHE ) SCHERING CORP.
XX
PI   Bond MW,  Campbell DS,  Lee ED,  Moore JG;
XX
DR   WPI; 1994-184511/22.
DR   P-PSDB; AAR53689.
XX
PI   New nucleic acid encoding new haematopoietic regulatory proteins
PT   - and related antibodies, useful for regulating progenitor cell
PT   differentiation and for diagnosis
XX
PS   Claim 1; Page 52; 59pp; English.
XX
CC   The sequence is that of the 38 kD haematopoietic regulatory protein,
CC   which is able to regulate differentiation of haematopoietic progenitor
CC   cells (HPCs), isolated by PCR using primers with sequences
CC   designed based on the N-terminal sequence of the isolated
CC   protein.
CC   See also AAQ66044-7.
XX
SQ   Sequence 2233 BP; 446 A; 650 C; 668 G; 489 T; 0 other;


alignment_scores:
              Quality:    9.00              Length:       9
                Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_75_83 x AAQ66045   ..

 Align seq 1/1  to: AAQ66045  from: 1  to: 2233

        1 AlaLysAlaProGlyAlaSerSerArg 9
          |||| ||||||||||.|||||||||||
      37 GCCAAAGCCCCCGGTGCCTCGTCGCGG 63

seq_name: ; ID: ; locality/genus/species/genotype; NA: 63 ; AC: AAC83266

seq_documentation_block:
ID   AAC83266 standard; DNA; 3317 BP.
XX
AC   AAC83266;
XX
DT   21-NOV-1994  (first entry)
XX
DE   Gland specific Ets transcription factor (GSEF) cDNA sequence.
XX
KW   Transcription factor; gland-specific Ets transcription factor; GSEF;
KW   metastatic potential; cancer; tumour; metastasis; breast; prostate;
KW   leukaemia; lymphoma; sarcoma; melanoma; chromosome 6p21.1-6p21.3; ss.
XX
OS   Homo sapiens.
XX
PN   WO2000070092-A1.
XX
PD   23-NOV-2000.
XX
PF   12-MAY-2000;  2000WO-US13173.
XX
PR   14-MAY-1999;   99US-0134112.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Kaufmann J,  Xin H,  Harrowe G;
XX
DR   WPI; 2001-041019/05.
DR   P-PSDB; AAB49633.
XX
PT   Detecting metastatic and potential metastatic cancerous cells, useful
PT   for diagnosing, prognosing, grading and staging of cancers by detecting
PT   gland specific Ets transcription factor gene product in a biological
PT   sample from a cell
XX
PS   Disclosure; Fig 2; 95pp; English.
XX
CC   This invention relates to a method for the detection and determination of
CC   the metastatic potential of a cell. The method comprises detecting a
CC   gland-specific Ets transcription factor (GSEF) gene product in a test
CC   sample. Detection of a GSEF gene product in the test sample in amount
```

CC lower than that in a normal cell, is indicative of a cell with high
CC metastatic potential. The method is useful for determining the metastatic
CC potential of a cell, for the diagnosis and prognosis of cancer as well as
CC grading and staging of cancers by detecting GSEF expression in a
CC biological test sample. The method may also be used to monitor patients
CC having a predisposition to develop a particular cancer. GSEF polypeptides
CC are useful for producing antibodies, in cancer diagnosis, prognosis,
CC grading, staging and management of breast and prostate tumours, and in
CC detecting polymorphisms in the sequence. GSEF genes and proteins are also
CC useful in gene therapy. GSEF gene product expression levels can be used
CC in conjunction with any tissue in which an alteration in GSEF gene
CC product expression levels is associated with development of a
CC cancer associated phenotype. Cancers, which can be monitored include
CC cancers of the prostate, cervix, lung and colon, melanomas, colorectal
CC adenocarcinomas, Wilm's tumour, retinoblastoma, sarcomas, myosarcomas,
CC lung carcinomas, leukaemia, and lymphomas. The GSEF gene is located on
CC human chromosome 6, specifically at 6p21.1 6p21.3. The present sequence
CC represents the cDNA encoding GSEF.
XX
SQ Sequence 3317 BP; 710 A; 1026 C; 970 G; 611 T; 0 other;

alignment scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2-COPY_75_83 x AAC83266 ..

Align seq 1/1 to: AAC83266 from: 1 to: 3317

1 AlaLysAlaArgGlyAlaGlnSerArg 9
||||||||||||||||||||||||||
2037 GCCAAGGCCCCTGGGGCCAGCAGCCGC 2063

Seq name: /SIDS3/gcgData/genseq_g/genseq_pg/2A1-COPY.AA...

seq documentation block:
ID AAZ23193 standard; DNA; 8805 BP.
XX
AC AAZ23193;
XX
DT 24-JAN-2000 (first entry)
XX
DE Mouse mammary tumor virus (MMTV) proviral genome nucleotide sequence.
XX
KW Immune response; mouse mammary tumor virus; MMTV; immunomodulatory;
KW epitope; passive immunotherapy; ss.
XX
OS Mouse mammary tumor virus.
XX
PN WO9951268-A1.
XX
PD 14-OCT-1999.
XX
PF 08-APR-1999; 99WO-US007712.
XX
PR 08-APR-1998; 98US-0057004.
XX
PA (UIBI) INT BIOIMMUNE SYSTEMS INC.
XX
PI Stewart THM, Gershoni J;
XX
DR WPI; 1999-620171/54.
DR P-PSDB; AAY41139, AAY41140, AAY41141, AAY41142, AAY41143.
XX
PT Novel immunomodulatory compositions of mouse mammary tumor virus (MMTV)
PT antigens used for treating or preventing MMTV infection.
XX
PS Disclosure; Fig 1; 92pp; English.
XX
CC The invention provides a method for inducing an immune response to mouse
CC mammary tumor virus (MMTV) in a human subject. The method comprises
CC administering to the subject an immunomodulatory composition comprising a
CC pharmaceutical carrier and at least one MMTV antigen (or a discontinuous
CC epitope of MMTV) in an amount sufficient to elicit an increase in a MMTV
CC specific cellular or humoral response. The compositions and methods of
CC the invention are useful for treating or preventing a human disease or
CC disorder caused by MMTV, or a biological or chemical agent that reacts
CC with an antibody directed towards an MMTV antigen. The compositions and
CC antibodies may be used in passive immunotherapy. The immunomodulatory
CC compositions can be used in a preventative manner for those subjects not
CC exposed to MMTV or MMTV antigen reactive agents. The MMTV antigens also
CC have use in diagnostic assays. The present sequence represents the
CC nucleotide sequence of the complete proviral genome of one strain of
CC MMTV.
XX
SQ Sequence 8805 BP; 2646 A; 1875 C; 1969 G; 2315 T; 0 other;

alignment scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2-COPY_75_83 x AAZ23193 ..

Align seq 1/1 to: AAZ23193 from: 1 to: 8805

1 AlaLysAlaArgGlyAlaSerSerArg 9
||||||||||||||||||||||||||
8749 GCCAAAGCCCCCTGGGGCCAGCAGCCGC 8775

Seq name: /SIDS3/gcgData/genseq_g/genseq_pg/NA1993.DAT:AAQ59397

seq documentation block:
ID AAQ59397 standard; cDNA; 282 BP.
XX
AC AAQ59397;
XX
DT 16-MAR-1994 (first entry)
XX
DE Human brain Expressed Sequence Tag EST00420.
XX
KW Gene transcription product; genetic markers; tagging; in vivo;
KW transcription; mapping; locations; chromosomes; chromosomal; ss.
XX
OS Homo sapiens.
XX
PN WO9315178-A.
XX
PD 19-AUG-1993.
XX
PF 12-FEB-1993; 93WO-US001294.
XX
PR 12-FEB-1992; 92US-0834195.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
XX
PI Adams MD; Moreno RF; Venter CJ;
XX
DR WPI; 1993-272882/34.
XX
PT Enriched oligonucleotides and corresp. sequences - used as
PT markers for human genes transcribed in vivo, facilitate tagging
PT of most human genes.
XX
PS Example 4; Page 168; 500pp; English.
XX
CC The Expressed Sequence Tag was isolated from a human brain cDNA
CC library as part of a large set of ESTs which can be used as markers
CC for human genes transcribed in vivo. They can be used to facilitate
CC tagging of most human genes, for mapping locations of expressed genes
CC on chromosomes, for individual or forensic identification, for mapping

```
CC   locations of disease-associated genes, for identification of tissue
CC   type, and for prepn. of antisense sequences, probes and constructs.
CC   EST00420 has a "poor" coding probability as evaluated using the
CC   coding-region prediction program CRM. See also AAQ59041-Q61440.
XX
SQ   Sequence 282 BP; 56 A; 71 C; 100 G; 50 T; 5 other;


alignment_scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_75_83 x AAQ59397/rev   ..

 Align seq 1/1  to reverse of: AAQ59397  from: 1  to: 282

        3 AlaProGlyAlaSerSerArg 9
          ||||||||||||||||||||
      244 GCTCCTGGAGCCTCCAGTAGA 224

seq_name: [illegible]

seq_documentation_block:
ID   AAF44809 standard; cDNA; 845 BP.
XX
AC   AAF44809;
XX
DT   27-MAR-2001  (first entry)
XX
DE   Cysteine protease coding sequence #23.
XX
KW   Cell death modulator; programmed cell death; PCD; apoptosis;
KW   forestry plant; ss.
XX
OS   Eucalyptus grandis.
XX
PN   WO200075331-A1.
XX
PD   14-DEC-2000.
XX
PF   02-JUN-2000; 2000WO-NZ00086.
XX
PR   04-JUN-1999;   99US-0325932.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Flinn B,  Lasham A;
XX
DR   WPI; 2001-061724/07.
DR   P-PSDB; AAB65782.
XX
PT   Novel defender against cell death polynucleotide useful for modulating
PT   programmed cell death pathway and specific development pathways in
PT   forestry plant -
XX
PS   Claim 1; Page 92; 142pp; English.
XX
CC   The present invention relates to coding sequences (see AAF44740-F44840
CC   and AAF44843-F44844) and proteins (see AAB65711-B65814) involved in
CC   programmed cell death (PCD; apoptosis). The coding sequences and proteins
CC   of the present invention are useful for modulating a PCD or cell death
CC   pathway and various developmental pathways in a forestry plant, by
CC   stably incorporating one of the present coding sequences into the genome
CC   of the forestry plant, where the coding sequence provides a PCD pathway
CC   that is not present in a native form of the forestry plant.
XX
SQ   Sequence 845 BP; 168 A; 244 C; 237 G; 186 T; 0 other;


alignment_scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_75_83 x AAF44809/rev   ..

 Align seq 1/1  to reverse of: AAF44809  from: 1  to: 845

        3 AlaProGlyAlaSerSerArg 9
          ||||||||||||||||||||
      289 GCTCCTGGAGCCTCCAGTAGA 269

seq_name: [illegible]

seq_documentation_block:
ID   AA235674 standard; DNA; 1173 BP
XX
AC   AA235674;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana [illegible] DNA fragment SEQ ID NO: 11022.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
```

PR 04 JUN 1999; 9908 0137502.
PR 07 JUN 1999; 9908 0137724.
PR 08 JUN 1999; 9908 0138094.
PR 10 JUN 1999; 9908 0138540.
PR 10 JUN 1999; 9908 0138847.
PR 14 JUN 1999; 9908 0139119.
PR 16 JUN 1999; 9908 0139452.
PR 16 JUN 1999; 9908 0139454.
PR 17 JUN 1999; 9908 0139492.
PR 18 JUN 1999; 9908 0139454.
PR 18 JUN 1999; 9908-0139455.
PR 18 JUN 1999; 9908 0139456.
PR 18 JUN 1999; 9908 0139457.
PR 18 JUN 1999; 9908 0139458.
PR 18 JUN 1999; 9908 0139459.
PR 18 JUN 1999; 9908 0139460.
PR 18 JUN 1999; 9908-0139461.
PR 18 JUN 1999; 9908 0139462.
PR 18 JUN 1999; 9908 0139464.
PR 18 JUN 1999; 9908-0139750.
PR 18 JUN 1999; 9908-0139764.
PR 21 JUN 1999; 9908 0139817.
PR 22 JUN 1999; 9908 0139899.
PR 23 JUN 1999; 9908 0140354.
PR 23 JUN 1999; 9908 0140354.
PR 24 JUN 1999; 9908 0140695.
PR 28 JUN 1999; 9908 0140823.
PR 29 JUN 1999; 9908 0140991.
PR 30 JUN 1999; 9908 0141287.
PR 01 JUL 1999; 9908 0141842.
PR 01 JUL 1999; 9908 0142154.
PR 02 JUL 1999; 9908 0142055.
PR 06 JUL 1999; 9908 0142390.
PR 08 JUL 1999; 9908 0142804.
PR 09 JUL 1999; 9908 0142920.
PR 12 JUL 1999; 9908 0142977.
PR 13 JUL 1999; 9908 0143542.
PR 14 JUL 1999; 9908 0143624.
PR 15 JUL 1999; 9908 0144005.
PR 16 JUL 1999; 9908 0144085.
PR 16 JUL 1999; 9908 0144086.
PR 19 JUL 1999; 9908 0144325.
PR 19 JUL 1999; 9908 0144341.
PR 19 JUL 1999; 9908 0144342.
PR 19 JUL 1999; 9908-0144344.
PR 19 JUL 1999; 9908-0144344.
PR 19 JUL 1999; 9908 0144345.
PR 20 JUL 1999; 9908-0144352.
PR 20 JUL 1999; 9908-0144642.
PR 20 JUL 1999; 9908 0144884.
PR 21 JUL 1999; 9908-0144814.
PR 21 JUL 1999; 9908 0145086.
PR 21 JUL 1999; 9908-0145088.
PR 22 JUL 1999; 9908-0145085.
PR 22 JUL 1999; 9908-0145087.
PR 22 JUL 1999; 9908-0145089.
PR 22 JUL 1999; 9908 0145192.
PR 23 JUL 1999; 9908 0145145.
PR 23 JUL 1999; 9908 0145218.
PR 23 JUL 1999; 9908 0145224.
PR 26 JUL 1999; 9908 0145276.
PR 27 JUL 1999; 9908 0145914.
PR 27 JUL 1999; 9908 0145918.
PR 27 JUL 1999; 9908 0145919.
PR 28 JUL 1999; 9908 0145951.
PR 02 AUG 1999; 9908 0146386.
PR 02 AUG 1999; 9908 0146388.
PR 02 AUG 1999; 9908 0146389.
PR 03 AUG 1999; 9908 0147048.
PR 04 AUG 1999; 9908-0147204.
PR 04 AUG 1999; 9908 0147302.
PR 05 AUG 1999; 9908 0147192.
PR 05 AUG 1999; 9908 0147260.
PR 06 AUG 1999; 9908 0147303.
PR 06 AUG 1999; 9908 0147416.
PR 09 AUG 1999; 9908 0147494.
PR 09 AUG 1999; 9908 0147935.
PR 10 AUG 1999; 9908 0148171.
PR 11 AUG 1999; 9908 0148319.
PR 12 AUG 1999; 9908 0148341.
PR 13 AUG 1999; 9908 0148565.
PR 13 AUG 1999; 9908 0148684.
PR 16 AUG 1999; 9908 0149368.
PR 17 AUG 1999; 9908 0149175.
PR 18 AUG 1999; 9908 0149426.
PR 20 AUG 1999; 9908 0149722.
PR 20 AUG 1999; 9908 0149723.
PR 20 AUG 1999; 9908 0149929.
PR 23 AUG 1999; 9908 0149902.
PR 24 AUG 1999; 9908 0149930.
PR 25 AUG 1999; 9908-0150566.
PR 26 AUG 1999; 9908 0150884.
PR 27 AUG 1999; 9908 0151065.
PR 27 AUG 1999; 9908 0151066.
PR 27 AUG 1999; 9908-0151080.
PR 30 AUG 1999; 9908 0151403.
PR 31 AUG 1999; 9908 0151438.
PR 01 SEP 1999; 9908 0151930.
PR 07 SEP 1999; 9908 0152463.
PR 10 SEP 1999; 9908 0153070.
PR 13 SEP 1999; 9908 0153758.
PR 15 SEP 1999; 9908 0154018.
PR 16 SEP 1999; 9908 0154039.
PR 20 SEP 1999; 9908 0154779.
PR 22 SEP 1999; 9908 0155139.
PR 23 SEP 1999; 9908 0155486.
PR 24 SEP 1999; 9908 0155659.
PR 28 SEP 1999; 9908 0156458.
PR 29 SEP 1999; 9908 0156596.
PR 04 OCT 1999; 9908 0157117.
PR 05 OCT 1999; 9908 0157754.
PR 06 OCT 1999; 9908 0157865.
PR 07 OCT 1999; 9908 0158029.
PR 08 OCT 1999; 9908 0158232.
PR 12 OCT 1999; 9908 0158869.
PR 13 OCT 1999; 9908 0159293.
PR 13 OCT 1999; 9908 0159294.
PR 13 OCT 1999; 9908 0159295.
PR 14 OCT 1999; 9908 0159329.
PR 14 OCT 1999; 9908 0159330.
PR 14 OCT 1999; 9908 0159331.
PR 14 OCT 1999; 9908 0159637.
PR 14 OCT 1999; 9908-0159638.
PR 18 OCT 1999; 9908 0159584.
PR 21 OCT 1999; 9908 0160741.
PR 21 OCT 1999; 9908 0160767.
PR 21 OCT 1999; 9908 0160768.
PR 21 OCT 1999; 9908-0160770.
PR 21 OCT 1999; 9908 0160814.
PR 21 OCT 1999; 9908 0160815.
PR 22 OCT 1999; 9908 0160980.
PR 22 OCT 1999; 9908 0160981.
PR 22 OCT 1999; 9908 0160989.
PR 25 OCT 1999; 9908 0161404.
PR 25 OCT 1999; 9908 0161405.
PR 25 OCT 1999; 9908 0161406.
PR 26 OCT 1999; 9908 0161359.
PR 26 OCT 1999; 9908 0161360.
PR 26 OCT 1999; 9908 0161361.
PR 28 OCT 1999; 9908 0161920.
PR 28 OCT 1999; 9908 0161992.
PR 28 OCT 1999; 9908 0161993.
PR 29 OCT 1999; 9908 0162142.

alignment scores:
Quality: 7.00 Length: 7

```
            Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_75_83 x AAC35674/rev

 Align seq 1/1  to reverse of: AAC35674  from: 1  to: 1173

        3 AlaProGlyAlaSerSerArg 9
          ||||||||||||||||||||
      202 GCTCCAGGAGCCAGCTCGAGG 182

seq_type: DNA/triplets/sequence_type: p_NA/ref: AAC AA_177/

seq_documentation_block:
ID   AAQ47726 standard; cDNA; 1292 BP.
XX
AC   AAQ47726;
XX
DT   21-FEB-1994  (first entry)
XX
DE   Coding sequence of NF-IL6 Beta.
XX
KW   Interleukin; hematinic; expression plasmid; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             89..898
FT                   /*tag= a
FT                   /product= "NF-IL6 Beta."
XX
PN   JP05192152-A.
XX
PD   03-AUG-1993.
XX
PF   27-NOV-1991;   91JP-0312667.
XX
PR   27-NOV-1991;   91JP-0312667.
XX
PA   (KISH/) KISHIMOTO T.
XX
DR   WPI; 1993-277468/35.
DR   P-PSDB; AAR40801.
XX
PT   Expressed recombinant NF-IL6beta gene - useful as hematinic.
XX
PS   Claim 1; Page 9-11; 11pp; Japanese.
XX
CC   Expressed recombinant NF-IL6 Beta is useful as a hematinic.  NF-IL6
CC   Beta was obtained by screening a lambda gt11 expression library
CC   with an antibody prepared by combining a synthetic peptide
CC   corresponding to 17 amino acids of NF-IL6 with ovalbumin and
CC   immunising a rabbit to give antiserum.  NF-IL6 Beta was used to
CC   construct two expression plasmids: CMV-NF-IL6 Beta and BS-NF-IL6
CC   Beta. The two plasmids were used to transform E. coli JM105.
XX
SQ   Sequence 1292 BP; 252 A; 447 C; 436 G; 287 T; 0 other.


alignment_scores:
            Quality:    7.00              Length:       7
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_75_83 x AAQ47726  ..

 Align seq 1/1  to: AAQ47726  from: 1  to: 1292

        3 AlaProGlyAlaSerSerArg 9
          ||||||||||||||||||||
      476 GCTCCAGGAGCCAGCTCGAGG 496
```

```
seq_type: DNA/triplets/sequence_type: p_NA/ref: AAC AA_59248

seq_documentation_block:
ID   AAC59248 standard; cDNA; 1984 BP.
XX
AC   AAC59248;
XX
DT   29-JAN-2001  (first entry)
XX
DE   Human secreted protein cDNA sequence #34.
XX
KW   [illegible] expression; secreted; ss.  Utopic; neuroprotective; antiviral;
KW   antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW   vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW   cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; human; secreted protein; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055199-A1.
XX
PD   21-SEP-2000.
XX
PF   09-MAR-2000; 2000WO-US06014.
XX
PR   12-MAR-1999;   99US-0124095.
PR   11-JUN-1999;   99US-0138598.
PR   03-DEC-1999;   99US-0168665.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA;  Kaiben SM;  Komatsoulis G;
XX
DR   WPI; 2000-572559/53.
DR   P-PSDB; AAB27827.
XX
PT   Isolated nucleic acid molecule encoding a human secreted protein is
PT   used in preventing, treating or ameliorating a medical condition
XX
PS   Claim 1; Page 360-361; 433pp; English.
XX
CC   The invention relate to the isolation of genes AAC59215-C59261 encoding
CC   47 human secreted proteins AAB27794-B27840. The genes can be used to
CC   generate fusion proteins by linking to the gene for the human
CC   immunoglobulin G Fc portion for increasing the stability of
CC   the fusion protein as compared to the human protein only. The genes and
CC   proteins are useful for preventing, ameliorating or treating medical
CC   conditions, e.g. by protein or gene therapy. The genes are isolated
CC   from a range of human tissues disclosed in the specification. The
CC   nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC   the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer, and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis, (c) cardiovascular disorders such as myocardial ischaemia, (d)
CC   wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
XX
SQ   Sequence 1984 BP; 650 A; 320 C; 346 G; 668 T; 0 other.


alignment_scores:
            Quality:    7.00              Length:       7
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_75_83 x AAC59218/rev  ..
```

```
Align_seq_1/1  to reverse of: AAC52248  from: 1  to: 1964

       2 LysAlaProGlyAlaSerSer 8
         ||||||||||||||||||||
    1413 AAGGCTCCTGGTGCAICTTCA 1393

Seq_name: "/tmp/3_relay/pens_pipeseq/pens_seq_NA2000.DAT:AAF25173

Seq_documentation_block:
ID   AAF25173 standard; cDNA; 1986 BP.
XX
AC   AAF25173;
XX
DT   30-APR-2001  (first entry)
XX
DE   Nucleotide sequence of a human protein having a hydrophobic domain.
XX
KW   human; hydrophobic protein; secretory protein; membrane protein; cytol;
KW   tumour inhibition; immune deficiency; autoimmune disorder; anaemia; burn;
KW   infectious disease; cancer; ulcer; periodontal disease; coagulation;
KW   Parkinson's disease; fertility; immune response; thrombosis; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             28..481
FT                   /*tag= a
FT                   /product= "hydrophobic protein"
XX
PN   WO200104297-A2.
XX
PD   18-JAN-2001.
XX
PF   16-JUN-2000; 2000WO-JP03942.
XX
PR   08-JUL-1999;  99JP-0194459.
XX
PA   (SAGA ) SAGAMI CHEM RES CENT.
PA   (PROT ) PROTEGENE INC.
XX
PI   Kato S,  Kimura T;
XX
DR   WPI; 2001-103081/11.
DR   P-PSDB; AAB41673.
XX
PT   Isolated human proteins and polynucleotides are used in research and
PT   have activities including cell proliferation/differentiation activity,
PT   immune stimulating activity and receptor/ligand activity
XX
PS   Claim 4; Page 135-137; 151pp; English.
XX
CC   The present sequence encodes a human protein with hydrophobic domains.
CC   AAF25163 represents a shorter version of the present sequence. The
CC   protein possesses a hydrophobic domain and so is a secretory protein
CC   or a membrane protein. The protein is used as an antigen to prepare
CC   antibodies. The polynucleotide sequence is useful as a source of probes
CC   for genetic diagnosis. It is also useful for producing the protein
CC   in large quantities and for gene therapy. The eukaryotic cells are used
CC   for detecting the receptors or ligands corresponding to the protein and
CC   for detecting small novel pharmaceuticals. The antibodies are also used
CC   for detection, quantification and purification of the proteins. Both the
CC   protein and polynucleotide may be used in research or as nutritional
CC   sources or supplements. The protein may have cytokine and cell
CC   proliferation/differentiation activity, immunostimulating or suppressive
CC   activity, haematopoiesis regulating activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity, homeostatic
CC   and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity and tumour inhibition activity. It may therefore may be used to
CC   treat immune deficiencies resulting from autoimmune disorders or
CC   infectious diseases, cancer, sepsis, anaemias, burns and ulcers,
CC   periodontal disease, Parkinson's disease, induce fertility, improve
CC   immune response and enhance coagulation or inhibit thrombosis.
XX
SQ   Sequence 1986 BP; 455 A; 620 C; 546 G; 417 T; 0 other;


alignment scores:
                Quality:    7.00              Length:       7
                  Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2 (COPY 75-83) x AAF25173/rev     .

 Align_seq_1/1  to reverse of: AAF25173  from: 1  to: 1986

       2 LysAlaProGlyAlaSerSer 8
         |||| ||||||||||||||||
    1578 AAGGCTCCTGGTGCAICTCCA 1558

Seq_name: "/tmp/3_relay/pens_pipeseq/pens_seq_NA2000.DAT:AAF15989

Seq_documentation_block:
ID   AAF15989 standard; cDNA; 3118 BP.
XX
AC   AAF15989;
XX
DT   12-MAR-2001  (first entry)
XX
DE   Human prostate cancer antigen nucleotide sequence SEQ ID NO:424.
XX
KW   Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW   neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
KW   vulnerary; gastrointestinal; nephrotropic; antiinfective; gynaecological;
KW   antibacterial; gene therapy; neural; immune; reproductive; renal;
KW   gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW   wound; infectious disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200053774-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05988.
XX
PR   12-MAR-1999;  99US-0124270.
XX
PA   (HUMA ) HUMAN GENOME SCI INC.
PA   (ROSE/) ROSEN C A.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-587913/55.
DR   P-PSDB; AAB56786.
XX
PT   Prostate cancer associated gene sequences, referred to as prostate
PT   cancer antigens, useful for treatment, prevention, and diagnosis of
PT   disorders such as prostate cancer
XX
PS   Claim 1; Page 906; 2338pp; English.
XX
CC   AAF15566 to AAF16505 encode the human prostate cancer associated
CC   proteins, called prostate cancer antigens, given in AAB56368 to AAB57402.
CC   The prostate cancer antigens can have neuroprotective, cytostatic,
CC   cardioactive, immunomodulatory, muscular, vulnerary, gastrointestinal,
CC   nephrotropic, antiinfective, gynaecological and antibacterial activities,
CC   and can be used in gene therapy. The prostate cancer antigen
CC   polynucleotides may be used for detection of prostate cancer, chromosome
CC   identification, as chromosome markers, and for numerous other diagnostic
CC   or research purposes. The prostate cancer antigens may be used to treat
CC   disorders such as neural, immune, muscular, reproductive,
CC   gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC   disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC   AAB57303 represent sequences used in the exemplification of the present
CC   invention.
```

```
XX
SQ   Sequence 3118 BP; 938 A; 636 C; 638 G; 902 T; 4 other;


alignment_scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_75_83 x AAF15989  ..

 Align seq 1/1  to: AAF15989  from: 1  to: 3118

       2 LysAlaProGlyAlaSerSer 8
         ||||||||||||||||||||
    1684 AAGGCTCCTGGTGCATCTTCA 1704

seq_score  "%5f%5l %score1ty/1e-nosq/genes-seq_NA.*%1 EAS AAF%f%%

seq_documentation_block:
ID   AAF54902 standard; cDNA; 4632 BP
XX
AC   AAF54902;
XX
DT   15-MAY-2001  (first entry)
XX
DE   Human activity dependent neurotrophic factor (ADNF) III cDNA.
XX
KW   Activity dependent neurotrophic factor III; ADNF III; cancer; gastric
KW   neuronal cell death; foetal alcohol syndrome; proliferating cell;
KW   restenotic plaque; atherosclerosis; restenosis; retinal hemangioblastoma;
KW   benign prostatic hyperplasia cell; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             347..1093
FT                   /*tag= a
FT                   /product= "activity dependent neurotrophic factor
FT                   (ADNF) III"
FT                   /note= "the nucleotides encoding amino acids
FT                   546-578 are missing from the figure"
XX
PN   WO200109311-A2.
XX
PD   08-FEB-2001.
XX
PF   28-JUL-2000; 2000WO-US20742.
XX
PR   30-JUL-1999;   99US-0364609.
XX
PA   (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Gozes I,  Brenneman DE,  Zamostiano R,  Gelber E,  Pinhasov A;
PI   Bassan M;
XX
DR   WPI; 2001-159867/16.
DR   P-PSDB; AAB81987.
XX
PT   Inhibiting the growth of malignant or benign proliferating cells (e.g.
PT   breast cancer or psoriatic cells, respectively), by contacting the
PT   cells with an activity dependent neurotrophic factor III antisense
PT   oligonucleotide -
XX
PS   Example 1; Fig 7; 57pp; English.
XX
CC   The present sequence encodes a human activity dependent neurotrophic
CC   factor (ADNF) III. ADNF III is involved in preventing neuronal cell
CC   death, and foetal alcohol syndrome. The specification describes a method
CC   for inhibiting the growth of a pathologically proliferating cell. The
CC   method comprises contacting the cell with an ADNF III antisense
CC   oligonucleotide. The method is useful for inhibiting and detecting the
CC   growth of pathologically proliferating or malignant cells, e.g. breast
CC   cancer, neuroblastoma, ovarian cancer, endometrial cancer, prostate
CC   cancer, bladder cancer, lung cancer, oesophageal cancer, neuroendocrine
CC   cancer, brain cancer, colon cancer, testicular cancer, pancreatic cancer
CC   or leukaemia. The method may also be used to inhibit and/or detect the
CC   growth of benign proliferating cells, e.g. restenotic plaques in vascular
CC   smooth muscle (e.g. atherosclerosis or restenosis), benign prostatic
CC   hyperplasia cells, retinal hemangioblastomas or psoriatic cells. The
CC   method may be used to inhibit the growth of cancer cells ex vivo or
CC   in vivo.
XX
SQ   Sequence 4632 BP; 1375 A; 1029 C; 1004 G; 1224 T; 0 other;


alignment_scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_75_83 x AAF54902/rev  ..

 Align seq 1/1  to reverse of: AAF54902  from: 1  to: 4632

       3 AlaProGlyAlaSerSerArg 9
         |||||||||||||||||||||
      56 GCGCCCGGGGCCTCGTCGCGC 36

seq_score  "%5f%5l %score1ty/1e-nosq/genes-seq_NA2000.DAT:AAA59146

seq_documentation_block:
ID   AAA59146 standard; DNA; 19053 BP.
XX
AC   AAA59146;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Nucleotide sequence of the antibiotic Tel aviv gene cluster.
XX
KW   Polyketide; antibiotic Tel-Aviv; cell wall synthesis;
KW   lipid-disaccharide-pentapeptide; qinalvitis; ss.
XX
OS   Myxococcus xanthus.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (2955..601)
FT                   /*tag= a
FT                   /note= "encodes AAB02666"
FT   CDS             complement (4702..3116)
FT                   /*tag= b
FT                   /note= "encodes AAB07667"
FT   CDS             complement (5595..4720)
FT                   /*tag= c
FT                   /note= "encodes AAB07668"
FT   CDS             6290..6793
FT                   /*tag= d
FT                   /note= "encodes AAB07669"
FT   CDS             6870..7106
FT                   /*tag= e
FT                   /note= "encodes AAB07670"
FT   CDS             7119..8378
FT                   /*tag= f
FT                   /note= "encodes AAB07671"
FT   CDS             8404..9878
FT                   /*tag= g
FT                   /note= "encodes AAB07672"
FT   CDS             9886..9964
FT                   /*tag= h
FT                   /note= "encodes AAB07673"
FT   CDS             9797..10878
FT                   /*tag= i
FT                   /note= "encodes AAB07674"
```

```
FT   CDS             10909..11421
FT                   /*tag   j
FT                   /note  "encodes AA007675"
FT   CDS             11474..12897
FT                   /*tag   k
FT                   /note  "encodes AA007676"
FT   CDS             12938..14891
FT                   /*tag   l
FT                   /note  "encodes AA007677"
FT   CDS             14909..14898
FT                   /*tag   m
FT                   /note  "encodes AA007678"
FT   CDS             14963..16213
FT                   /*tag   n
FT                   /transl_except  (pos: 15521..15524, aa: Lys)
FT                   /note  "encodes AA007679"
FT   CDS             16224..17009
FT                   /*tag   o
FT                   /transl_except  (pos: 16563..16595, aa: Ser)
FT                   /note  "encodes AA007680"
FT   CDS             17000..17767
FT                   /*tag   p
FT                   /note  "encodes AA007681"
FT   CDS             17782..19053
FT                   /*tag   q
FT                   /note  "encodes AA007682"
XX
PN   EP1026248 A2.
XX
PD   09 AUG 2000.
XX
PF   31 JAN 2000; 2000EP-0400747.
XX
PR   29 JAN 1999;   99US-0240547.
XX
PA   (UYRA ) UNIV RAMOT APPLIED RES & IND DEV LTD.
XX
PI   Rosenberg E,  Ron E,  Ori E,  Paitan Y;
XX
DR   WPI; 2000-500254/45.
DR   P-PSDB; AAB07666, AAB07667, AAB07668, AAB07669, AAB07670, AAB07671,
DR   AAB07672, AAB07673, AAB07674, AAB07675, AAB07676, AAB07677, AAB07678,
DR   AAB07679, AAB07680, AAB07681, AAB07682.
XX
PT   Novel DNA sequence involved in polyketide antibiotic Tel Aviv
PT   production useful for inhibiting cell wall synthesis and in wide range
PT   of clinical applications such as treating gingivitis
XX
PS   Disclosure; Page 12-19; 66pp; English.
XX
CC   The specification describes a DNA sequence which partially encodes
CC   a functional portion of polypeptide component required for synthesizing
CC   the polyketide antibiotic Tel-Aviv, postmodification of antibiotic
CC   Tel aviv, or regulation of biosynthesis of antibiotic Tel-aviv. The
CC   antibiotic Tel aviv is a macrocyclic polyketide synthesised through
CC   the incorporation of acetate, methionine, and glycine. It inhibits cell
CC   wall synthesis by interfering with the polymerisation of the
CC   lipid disaccharide pentapeptide. Antibiotic Tel aviv genes are useful
CC   in combinatorial genetics, and for encoding protein components for the
CC   synthesis, modification and regulation of antibiotic antibiotic Tel aviv.
CC   Antibiotic Tel aviv is useful in a wide range of clinical applications
CC   such as treating gingivitis. Antibiotic Tel aviv is also useful for
CC   generating new biological agents from its secondary metabolites. The
CC   present sequence encodes proteins involved in synthesis of antibiotic
CC   Tel aviv.
XX
SQ   Sequence 19053 BP; 2947 A; 6220 C; 6640 G; 3245 T; 0 other;


alignment scores:
            Quality:     7.00              Length:       7
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000
```

alignment block:

Align seq 1/1 to reverse of: AAA59146 from: 1 to: 19053

```
        0 AlaProGlyAlaSerSerArg 9
          |||| |.|||||||||||||
    12298 GCTCCTGGGGCGTCATCAGGG 12278
```

```
OM of: US-09-126-945B-2_COPY_75_83 to: EST:*   out_format : pir

Date: Sep 14, 2001  7:47 PM

About: Results were produced by the GenCore software, version 4.5.
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL framer_p2n.model -DEV xlp
 -Q "/tmp_1/USPTO_sp_2_09/us-09-126-945b-2_copy_75_83.olip2n.rst_11***.*" -Ia "EST:*" -SUF "/gi_pir*.fasta" -T 1 -U 0
 -DB EST -QFMT fastap -SUFFIX olip2n.rst -GAPOP 4.500
 -GAPEXT 0.050 -MINMATCH 0.100 -LooPCL 0.000 -LooPEXT 0.000
 -QGAPOP 4.500 -QGAPEXT 0.050 -XGAPOP 60.000 -XGAPEXT 60.000
 -YGAPOP 6.000 -YGAPEXT 7.000 -YGAPOP 60.000 -YGAPEXT 60.000
 -DELOP 6.000 -DELEXT 7.000 -START 1 -MATRIX oligo
 -TRANS human40.cdi -LIST 45 -DOCALIGN 200 -THR_SCORE quality
 -THR_MIN 1 -ALIGN 15 -MODE LOCAL -OUTFMT pir  NORM ext  MINLEN 0
 -MAXLEN 2000000000  USER US09126945B_2COPY_75_83 -NCPU 4
 -ICPU 4 -LONGLOG -NO_XLPXY -WAIT -THREADS 1

Search information block:
Query: US-09-126-945B-2_copy_75_83
Query length: 9
Database: EST:*
Database sequences: 10228115
Database length: 4314594S4
Search time (sec): 9087.630000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000


score_list:
Sequence              Strd Orig    ZScore    EScore Len   ! Documentation  ..
qb_est97:BG179568      +     9.00   155.13     16.01    991  ! BG179568 6823281B2F1 NIH_MGC_91
qb_est70:BE071840      -     8.00   146.12     50.80    271  ! BE071840 RC2 BT0522 120200 014
qb_est84:BF116909      +     8.00   144.64     61.44    340  ! BF116909 nz03f03.y1 NCI_CGAP_M
qb_est89:BF548125      +     8.00   144.16     65.36    366  ! BF548125 UI-R-A0-bf-f-04-0-UI.
qb_est2:AA096594       -     8.00   142.96     76.27    440  ! AA096594 nm52c08.r1 Stratagene
qb_est97:BG114929      -     8.00   142.64     79.45    462  ! BG114929 af27e04.y1 Soares_pl
qb_est12:AA817715      -     8.00   142.50     80.89    472  ! AA817715 UI-R-A0-aa-e-10-0-UI.s
qb_est10:AA681554      +     8.00   142.28     83.19    489  ! AA681554 vr28a04.r1 Barstead
qb_qss4:CNS04392       -     8.00   137.15    160.52   1069  ! AL272465 Tetraodon oligo-fridl
qb_hte:AK005710        +     8.00   134.65    212.87   1497  ! AK005710 Mus musculus adult ma
qb_qss14:AQ643973      -     7.00   133.60    253.24    127  ! AQ643973 RPCI93-8pnI1-80P6.I1.R
qb_est46:AW485944      +     7.00   133.35    261.57    132  ! AW485944 CM0-LT0069-140299-144-
qb_est42:AW062078      +     7.00   133.20    266.55    135  ! AW062078 660009D12.y1 660 - Mix
qb_est47:AW434511      -     7.00   133.06    271.51    138  ! AW434511 UI-R-BJ0-pdf-a-02-0-U
qb_qss34:AZ918723      -     7.00   133.01    273.15    139  ! AZ918723 1006004D11.2EL_x1 1006
qb_est51:AW751866      +     7.00   131.89    315.38    165  ! AW751866 QV2-CT0121-280999-001-
qb_est14:AA963296      -     7.00   131.81    318.58    167  ! AA963296 UI-R-E1-qi-a-09-0-UI.s
qb_est50:AW714310      +     7.00   131.62    326.56    172  ! AW714310 h7d03ne.f1 Neurospora
qb_est92:BF747211      -     7.00   131.54    329.74    174  ! BF747211 RC4 BT0503-241100-013
qb_est50:AW712067      +     7.00   131.04    351.84    188  ! AW712067 f8b02ne.f1 Neurospora
qb_est50:AW662944      +     7.00   130.80    362.79    195  ! AW662944 hh94b11.y1 NCI_CGAP_GI
qb_est21:AI547568      -     7.00   130.73    365.91    197  ! AI547568 UI-R-C3-sr-f-09-0-UI.s
qb_est108:T32165       +     7.00   130.19    392.20    214  ! T32165 EST44485 Human Brain Hom
qb_est74:BE425601      +     7.00   129.92    405.98    223  ! BE425601 WHP0322_Au3_Au42S Whe
qb_est87:BF419143      -     7.00   129.83    410.55    226  ! BF419143 UI-R-BJ2-bqc-f-02-0-UI
qb_est104:F32423       -     7.00   129.66    419.67    232  ! F32423 HSPD25156 HM3 Homo sapie
qb_est33:AV156415      -     7.00   129.44    431.77    240  ! AV156415 AV156415 Mus musculus
qb_est106:N75579       -     7.00   129.41    433.28    241  ! N75579 za85c03.s1 Soares_fetal_
qb_est47:AW452644      -     7.00   129.41    439.30    245  ! AW452644 UI-H-Bi3-alu-f-07-0-UI
qb_est21:AI511386      -     7.00   129.10    451.29    253  ! AI511386 UI-R-C3-st-q-01-0-UI.s
qb_est51:AW751860      +     7.00   129.10    451.29    253  ! AW751860 QV2-CT0121-280999-001-
qb_est105:H42182       -     7.00   129.04    454.25    255  ! H42182 yo11h02.r1 Soares_adult
qb_est19:AI382542      -     7.00   128.89    463.22    261  ! AI382542 qz03c09.x1 NCI_CGAP_Br
qb_est68:BB576415      +     7.00   128.89    463.22    261  ! BB576415 BB576415 RIKEN full-le
qb_est47:AW453069      -     7.00   128.84    466.20    263  ! AW453069 UI-H-BW1-amaeq-02-0-UI
qb_est92:BF746792      -     7.00   128.84    466.20    263  ! BF746792 RC4 BT0520-121100-012-
qb_est50:AW713448      +     7.00   128.67    476.59    270  ! AW713448 q8b10ne.f1 Neurospora
qb_est1:AA007573       -     7.00   128.57    482.49    274  ! AA007573 zh95h09.s1 Soares_fetu
qb_est105:H87872       -     7.00   128.55    483.99    275  ! H87872 yw16b02.r1 Soares_placen
qb_est13:AA612454      -     7.00   128.39    494.27    282  ! AA612454 nm52c04.s1 NCI_CGAP
qb_est106:M78272       -     7.00   128.39    494.27    282  ! M78272 EST00420 Fetal brain,
qb_est46:AW475158      -     7.00   128.25    504.07    288  ! AW475158 QV2-CT0121-280999-0
qb_est80:BE853100      -     7.00   128.23    504.53    289  ! BE853100 nw29e05.x1 Soares_t
qb_est87:BF402904      -     7.00   128.20    506.00    290  ! BF402904 UI-R-CA0-lhi-h-07-0
qb_est92:BF813470      -     7.00   128.09    514.30    295  ! BF813470 MR2-CT0127-271100-0

seq_name: qb_est97:BG179568

seq_documentation_block:
LOCUS        BG179568        991 bp    mRNA            EST       06-FEB-2001
DEFINITION   6823281B2F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4429413 5',
             mRNA sequence.
ACCESSION    BG179568
VERSION      BG179568.1  GI:12686271
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 991)
  AUTHORS    NIH MGC Project (https://mgc.nci.nih.gov).
  TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL    Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: NCI/CGAP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM10181  row: d  column: 22
             High quality sequence stop: 655.
FEATURES             Location/Qualifiers
     source          1..991
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4429413"
                     /clone_lib "NIH_MGC_91"
                     /tissue_type "adenocarcinoma, cell line"
                     /lab_host "DH10B (phage-resistant)"
                     /note "organ: prostate; Vector: pCMV SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.4 kb. Library enriched for
                     full length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      190 a    351 c    305 g    145 t
ORIGIN


alignment_scores:
            Quality:    9.00              Length:      9
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_75_83 x BG179568   ..

 Align Seq 1/1   to: BG179568  from: 1  to: 991
                 .
        1 AlaLysAlaProGlyAlaSerSerArg 9
          |||||||||||||||||||||||||||
      126 GCCAAGGCCCCTGGGGCAGCAGTCGG 152

seq_name: qb_est70:BE071840

seq_documentation_block:
LOCUS        BE071840        271 bp    mRNA            EST       09-JUN-2000
DEFINITION   RC2 BT0522 120200 014 c01 BT0522 Homo sapiens cDNA, mRNA sequence.
ACCESSION    BE071840
VERSION      BE071840.1  GI:9417280
```

```
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 271)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., de Oliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun Sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo SP,
            Brazil
            Tel: +55 11 2704922
            Fax: +55 11 2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/cgi-bin/db/dbget.pl?id=%7C,
            200 014 601&t3 2000 02 12&t4 1)
            Seq primer: puc 18 forward
            High quality sequence start: 22
            High quality sequence stop: 86.
FEATURES             Location/Qualifiers
     source          1..271
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="BM052"
                     /dev_stage="Adult"
                     /note="Organ: breast; Vector: puc18; Site 1: SmaI; Site 2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector, Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT       61 a      57 c      71 g      82 t
ORIGIN


alignment scores:
            Quality:     8.00              Length:       8
              Ratio:    1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2 COPY 75-83 x BE071840/rev     .

 Align seq 1/1  to reverse of: BE071840  from: 1  to: 271

        1 AlaLysAlaProGlyAlaSerSer 8
          ||||||||||||||||||||||||
      165 GCAAAAGCCCCAGGAGCTTCCTCA 142

seq name: qb_est80:BF116909

seq documentation block:
LOCUS       BF116909       340 bp    mRNA             EST       29-DEC-2000
DEFINITION  uz03103.y1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3667997 5'
            similar to SW:LY6E_MOUSE Q64129 LYMPHOCYTE ANTIGEN 6E PRECURSOR
            ;, mRNA sequence.
ACCESSION   BF116909
VERSION     BF116909.1  GI:10986685
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 340)
  AUTHORS   NCI-CGAP Mam Group, Gebster,J.A., Jacobs,M.D., Yuan,X., Kim,H.G.,
            Margaryan,N., Wawrzyniak,J.A., Berg,A., Bhat,N.K., Phillips,G.N.
            Jr., Haynes,M.K., Kennett,S.B., Miller,H., Braun,R.E., Hadjantonakis
            ,A.K., Cherry,J.L., Strausberg,R.L., Klapa,M.I. and Buetow,K.H.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            image.llnl.gov/image/html/iresources.shtml

            MGI:3428765
            Trace considered overall poor quality
            Seq primer:  40RP from Gibco
            High-quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1..340
                     /organism="Mus musculus"
                     /strain="C57/B6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3667997"
                     /clone_lib="NCI_CGAP_Mam5"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="7 months"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV SPORT6; Site 1: SalI;
                     Site 2: NotI; Cloned unidirectionally.  Primer: oligo dT
                     library constructed by Life Technologies.  Investigators
                     providing samples: Lothar Hennighausen/Robin Humphreys,
                     NIH"
BASE COUNT       109 a      84 c      62 g      85 t
ORIGIN


alignment scores:
            Quality:     8.00              Length:       8
              Ratio:    1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2 COPY 75-83 x BF116909     .

 Align seq 1/1  to: BF116909  from: 1  to: 340

        1 AlaLysAlaProGlyAlaSerSer 8
          |||| |||||||||||||||||||
      255 GCCAAAGCTCCGGGTGCCAGTTCA 278

seq name: qb_est80:BF548125

seq documentation block:
LOCUS       BF548125       366 bp    mRNA             EST       11-DEC-2000
DEFINITION  UI-R-A0-bt-1-04-0-UI.r1 UI-R-A0 Rattus norvegicus cDNA clone
            UI-R-A0-bt-1-04-0-UI-5', mRNA sequence.
ACCESSION   BF548125
VERSION     BF548125.1  GI:11639232
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 366)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
```

```
  JOURNAL   Genome Res. 6 (9), 791 806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            clones will be available through Research Genetics (www.resgen.com)
            This clone is also available through the I.M.A.G.E. Consortium at
            LLNL (info@image.llnl.gov). IMAGE ID  1793837
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .366
                     /organism "Rattus norvegicus"
                     /strain "Sprague Dawley"
                     /db_xref "taxon:10116"
                     /clone="UI-R-A0-bf-1-04-0 UI"
                     /clone_lib "UI-R-A0"
                     /dev_stage "adult"
                     /lab_host "DH10B (Life Technologies)"
                     /note="Vector: pT7T3D Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; This library
                     consists of a mixture of individually tagged normalized
                     libraries constructed from rat placenta, adult lung, brain
                     , liver, kidney, heart, spleen, ovary, and muscle. The tag
                     is a string of 3-5 nucleotides present between the Not I
                     site and the oligo dT track which allows identification of
                     the library of origin of a clone within the mixture."
BASE COUNT        71 a    114 c     92 g     89 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:       8
              Ratio:   1.000              Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_75_83 x BF548125   ..

 Align seq 1/1  to: BF548125  from: 1  to: 366

       2 LysAlaProGlyAlaSerSerArg 9
         |||||||||||||||||||||||||
     242 AAAGCACCCGGGGCTTCTTCAAGG 265

seq_name: qb_est2:AA096594

seq_documentation_block:
LOCUS       AA096594      440 bp    mRNA            EST       06-FEB-1997
DEFINITION  mm5.2c08.r1 Stratagene mouse melanoma (#937412) Mus musculus cDNA
            clone IMAGE:525134 5', mRNA sequence.
ACCESSION   AA096594
VERSION     AA096594.1  GI:1642496
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus
REFERENCE   1  (bases 1 to 440)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The Wash0-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            Wash0 HHMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:318982
            Seq primer:  28ml3 rev1 ET from Amersham
            High quality sequence stop: 440.
FEATURES             Location/Qualifiers
     source          1. .440
                     /organism "Mus musculus"
                     /db_xref "taxon:10090"
                     /clone "IMAGE:525134"
                     /clone_lib "Stratagene mouse melanoma (#937412)"
                     /tissue_type "melanoma"
                     /dev_stage "M2 cells"
                     /lab_host "SoloMax (kanamycin resistant)"
                     /note "Strain, Skin, Vector, pBluescript SK+; Site_1: EcoRI
                     ; Site_2: XhoI; Cloned unidirectionally.  Primer:  oligo
                     dT. From M2 cells, a highly metastatic derivative of the
                     K 1735 (mouse) melanoma. Average insert size: 1.6 Kb;
                     Uni-ZAP XR Vector;  5' adaptor sequence: 5' GAATTCGGCACGAG
                     3'  3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"
BASE COUNT        99 a    114 c    144 g     83 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:       8
              Ratio:   1.000              Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_75_83 x AA096594   ..

 Align seq 1/1  to: AA096594  from: 1  to: 440

       2 LysAlaProGlyAlaSerSerArg 9
         |||||||||||||||||||||||||
      86 AAGGCACCCGGGGCTTCTTCAAGG 109

seq_name: qb_est97:BG144926

seq_documentation_block:
LOCUS       BG144926      462 bp    mRNA            EST       01-FEB-2001
DEFINITION  qt79e04.y1 Soares_mouse_NMGB_bcell1 Mus musculus cDNA clone
            IMAGE:3333991 5' similar to TR:Q9VGU9 Q9VGU9 CG6608 PROTEIN. ;,
            mRNA sequence.
ACCESSION   BG144926
VERSION     BG144926.1  GI:12648330
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 462)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Other_ESTs: qt79e04.x1
            Contact: Robert Strausberg, Ph.D.
            Email: cgaphs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1078155
            Seq primer: -40RP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .462
                     /organism "Mus musculus"
                     /db_xref "taxon:10090"
                     /clone "IMAGE:3333991"
```

```
                     /clone_lib "Soares mouse NMGB bc II"
                     /lab_host "DH10B (phage resistant)"
                     /note "Organ: germinal B cell; Vector: pZF4D Fac
                     (Pharmacia) with a modified polylinker; Site 1: Not I;
                     Site 2: Eco RI; 1st strand cDNA was primed with a Not I
                     oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCTCGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pZF4D vector.
                     Library is normalized, constructed by Wendy Soares and
                     M.Fatima Bonaldo."
BASE COUNT       101 a     133 c     128 g      97 t       3 others
ORIGIN


alignment_scores:
            Quality:     8.00              Length:       8
              Ratio:    1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 HS_09_126_945B_2_COPY_75_83 x BG144926

 Align seq 1/1  to: BG144926  from: 1  to: 462

       2 LysAlaProGlyAlaSerSerArg 9
         ||||||||||||||||||||||||
     404 AAGGCACCCGGGGCTTCTTCAAGG 426

seq name: qb_est12:AA817715

seq documentation_block:
LOCUS        AA817715       472 bp     mRNA             EST       03-JUL-1999
DEFINITION   UI-R-A0-aa-e-10-0-UI.s1 UI-R-A0 Rattus norvegicus cDNA clone
             UI-R-A0-aa-e-10-0-UI 3' similar to gb|AA681553|AA681553 vr38a04.r1
             Barstead mouse myotubes MPLRB5 Mus musculus cDNA clone IMAGE:1122894 5',
             mRNA sequence.
ACCESSION    AA817715
VERSION      AA817715.1  GI:4198244
KEYWORDS     EST.
SOURCE       Norway rat.
  ORGANISM   Rattus norvegicus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae;
             Rattus.
REFERENCE    1  (bases 1 to 472)
  AUTHORS    Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE      Normalization and subtraction: two approaches to facilitate gene
             discovery
  JOURNAL    Genome Res. 6 (9), 791-806 (1996)
  MEDLINE    97044477
COMMENT      On Feb 17, 1998 this sequence version replaced gi:2887595
             Contact: Soares, MB
             Program for Rat Gene Discovery and Mapping
             University of Iowa
             451 Eckstein Medical Research Building Iowa City, IA 52242, USA
             Tel: 319 335 8250
             Fax: 319 335 9666
             Email: msoares@blue.weeg.uiowa.edu
             The sequence tag present in the cDNA between the NotI site and the
             oligo dT track served to identify it as a clone from the normalized
             adult ovary library. cDNA Library Preparation: M. Fatima Bonaldo,
             Ph.D. Clone distribution: clones will be available through Research
             Genetics This clone is also available through the I.M.A.G.E
             Consortium at LLNL (info@image.llnl.gov) IMAGE ID 1267892
             Seq primer: M13 Forward
             POLYA NO.
FEATURES             Location/Qualifiers
     source          1..472
                     /organism="Rattus norvegicus"
                     /strain "Sprague Dawley"
                     /db_xref "taxon:10116"
                     /clone "UI-R-A0-aa-e-10-0-UI"
                     /clone_lib "UI-R-A0"
                     /dev_stage "adult"
                     /lab_host "DH10B (Life Technologies)"
                     /note "Vector: pZF4D Fac (Pharmacia) with a modified
                     polylinker; Site 1: Not I; Site 2: Eco RI; This library
                     consists of a mixture of individually tagged normalized
                     libraries constructed from rat placenta, adult lung, brain
                     , liver, kidney, heart, spleen, ovary, and muscle. The tag
                     is a string of 3-5 nucleotides present between the Not I
                     site and the oligo dT track which allows identification of
                     the library of origin of a clone within the mixture."
BASE COUNT       127 a     130 c     111 g      124 t
ORIGIN


alignment_scores:
            Quality:     8.00              Length:       8
              Ratio:    1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 HS_09_126_945B_2_COPY_75_83 x AA817715/rev  ..

 Align seq 1/1  to reverse of: AA817715  from: 1  to: 472

       2 LysA aProGlyAlaSerSerArg 9
         |||| ||||||||||||||||||
     448 AAGGCACCCGGGGCTTCTTCAAGG 425

seq name: qb_est10:AA681553

seq documentation_block:
LOCUS        AA681553       488 bp     mRNA             EST       05-DEC-1997
DEFINITION   vr38a04.r1 Barstead mouse myotubes MPLRB5 Mus musculus cDNA clone
             IMAGE:1122894 5', mRNA sequence.
ACCESSION    AA681553
VERSION      AA681553.1  GI:2663693
KEYWORDS     EST.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 488)
  AUTHORS    Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Isaberpe,I.,
             Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
             Schellenberg,K., Stepton,M., Tan,F., Underwood,K., Moore,B.,
             Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
             Waterston,R.
  TITLE      The WashU-RHMI Mouse EST Project
  JOURNAL    Unpublished (1996)
COMMENT      Contact: Marra M/Mouse EST Project
             WashU RHMI Mouse EST Project
             Washington University School of Medicine
             4444 Forest Park Parkway, Box 8501, St. Louis, Mo 63108
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: mouseest@watson.wustl.edu
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             MGI:612230
             Seq primer:  28ml3 rev2 EF from Amersham
             High quality sequence stop: 487.
FEATURES             Location/Qualifiers
     source          1..488
                     /organism="Mus musculus"
                     /strain "C3H"
                     /db_xref "taxon:10090"
                     /clone "IMAGE:1122894"
                     /clone_lib "Barstead mouse myotubes MPLRB5"
                     /sex "male"
                     /lab_host "DH10B"
                     /note "Vector: pZF4D Fac (Pharmacia) with a modified
```

polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
was primed with a Not I oligo(dT) primer [5'
TGTTACGAATTCTGAAGTGGGAGGCGGCCGCTTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[AATTCGGATCCTTG], digested with Not I and cloned into the
Not I and Eco RI sites of the modified pF7F3 vector.
Library constructed by Bob Barstead. The C2C12 cell line
(available from ATCC, catalog # CRL-1772) differentiates
rapidly, forming contractile myotubes and producing
characteristic muscle proteins. "
BASE COUNT 110 a 127 c 154 g 97 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_75_83 x AA681553 ..

Align seq 1/1 to AA681553 from 1 to 499

2 LysAlaProGlyAlaSerSerArg 9
||||||||||||||||||||||||
18 AAGGCACCCGGGGCTTCTTCAAGG 41

seq_name: qb_qss4:CNS04392

seq_documentation_block:
LOCUS CNS04392 1069 bp DNA GSS 10-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence 17 end of clone
079M19 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL272495
VERSION AL272495.1 GI:7994582
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 1069)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE Charaterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1069)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1069)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) [illegible]
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetra.html
FEATURES Location/Qualifiers
source 1..1069
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="079M19"
/clone_lib="G"
/note="Genoscope sequence ID = CUR0079A01G1PID end = 17"

BASE COUNT 145 a 299 c 395 g 219 t 13 others
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_75_83 x CNS04392/rev ..

Align seq 1/1 to reverse of: CNS04392 from: 1 to: 1069

2 LysAlaProGlyAlaSerSerArg 9
||||||||||||||||||||||||
920 AAAGCCCCGGGAGCGTCCTCCCGG 897

seq_name: qb_hte:AK005710

seq_documentation_block:
LOCUS AK005710 1497 bp mRNA HTC 08-FEB-2001
DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched
library, clone:1700007G02, full insert sequence.
ACCESSION AK005710
VERSION AK005710.1 GI:12838426
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA,
clone_lib:RIKEN full-length enriched mouse cDNA library
clone:1700007G02.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE 1 (sites)
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High efficiency full-length cDNA cloning
JOURNAL Methods Enzymol. 303, 19-44 (1999)
REFERENCE 2 (sites)
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper selected cDNAs to
prepare full length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
REFERENCE 3 (sites)
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,I.,
Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,I.,
Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yamaga,Y., Ishikawa,T., Ozawa,K., Tanaka,I.,
Matsuura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20540914
REFERENCE 4 (sites)
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and
FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5 (bases 1 to 1497)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
Hamaguchi,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
Ishii,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
Kojima,Y., Konno,H., Kondo,S., Koya,S., Kurihara,C., Matsuyama,I.,
Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Oshima,M., Ota,M.,
Ohsato,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Taguchi,M., Tagawa,A., Takahashi,F.,
Tanaka,I., Tejima,Y., Toya,T., Yamamura,I., Yasunishi,A.,
Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.

```
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN,
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo activated reverse transcriptase
            and subsequently enriched for full length by cap trapper. Second
            strand cDNA was prepared with the primer adapter of sequence[5'
            GAGAGAGAGAGGGCCGCAATTAATTCTCGAGTTAATTAAATTAATCCCCCCCCCC 3'], cDNA
            was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3'
            end: SstI. Host: SOLR.
FEATURES             Location/Qualifiers
     source          1..1497
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /db_xref="MGD:MGI:1890440"
                     /db_xref="MGD:MGI:1914533"
                     /clone="1700007G02"
                     /sex="male"
                     /tissue_type="testis"
                     /clone_lib="RIKEN full length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             115..1071
                     /note="putative"
                     /codon_start=1
                     /protein_id="BAB24199.1"
                     /db_xref="GI:12838427"
                     /translation="MVGYDAKADVEGNSKEEVAVAGSVSLVTKAESESEEVSEEEGG
                     LQTERLCPSDPNAEYHGIEQAAEQLLQEFCPRAFWFCRVDAQTLSDCYCAVQELAEES
                     LTELLYQANLYQTHQFSAHEVCGGLSAGTATLTVHPVDVLRTRLAAQGEPKIYNNLRE
                     AIRTMYEEEGPEVFYKGLIPCVIAIFPYAGLQFSVYEYLKSVYSGKLLDETQQVLQEE
                     NLLCGCGSGVISKEEIYPLDLEEERLQVGGFLHAGSAELQVRSYRGLLDETQQVLQEE
                     GTRGFEKGLSPSLMKAALSTGEMFFWYELFCNLFHCIRREDR"
BASE COUNT       346 a     416 c     417 g     318 t
ORIGIN


alignment scores:
            Quality:     8.00              Length:       8
              Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2-COPY_75_83 x AK005710  ..

 Align seq 1/1  to: AK005710  from: 1  to: 1497

        2 LysAlaProGlyAlaSerSerArg 9
          ||||||||||||||||||||||||
      941 AAGGCACCCGGGGCTTCTTCAAGG 964

seq_name: gb_gss14:AQ643973

seq_documentation_block:
LOCUS        AQ643973       127 bp    DNA              GSS       08-JUL-1999
DEFINITION  RPCI93-bpnII-30D9.TJ RPCI93 bpnII Trypanosoma brucei genomic clone
            RPCI93-bpnII-30D9, DNA sequence.
ACCESSION   AQ643973
VERSION     AQ643973.1  GI:5120683
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei.
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 127)
  AUTHORS   El-Sayed,N., Zhao,B., Gill,S., Suh,E., Malek,J., Fujii,C.,
            Gerrard,C., Leech,V., de Jong,P., Hillu,E., Melville,S., Donelson,J.
            , Fraser,C. and Adams,M.
  TITLE     Use of BAC end sequences from Trypanosoma brucei GUTat 10.1 RPCI 93
            ...clear, for gene discovery and sequence ready map construction
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Najib M. El-Sayed
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: nelsayed@tigr.org
            Clones and high density filters may be purchased from BACPAC
            Resources (http://bacpac.med.buffalo.edu). BAC end sequences search
            page: http://www.tigr.org/tdb/mdb/trbgi/.
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..127
                     /organism="Trypanosoma brucei"
                     /strain="GUTat 10.1"
                     /db_xref="taxon:5691"
                     /clone="RPCI93-bpnII-30D9"
                     /clone_lib="RPCI93-bpnII"
                     /note="Vector: pBACe3.6; Site 1: BamHI; Site 2: BamHI;
                     Constructed for The Institute for Genomic Research by
                     Bohui Zhao in Pieter de Jong's laboratory (Roswell Park
                     Cancer Institute, Buffalo, NY). Briefly, Trypanosoma
                     brucei TREU927/4 GUTat 10.1 agarose embedded DNA was
                     partially digested with a combination of Eco RI and Eco RI
                     methylase (RPCI93-EcoRI segment) or Dpn II (RPCI93-bpnII
                     segment). High molecular weight fragments were ligated in
                     pBACe3.6 vector digested with Eco RI or BamHI,
                     respectively. The average insert size is 141 Kb. Total
                     coverage (both segments): ~0.90 x the haploid
                     size of minichromosomal genome."
BASE COUNT        30 a      29 c      35 g      33 t
ORIGIN


alignment scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2-COPY_75_83 x AQ643973/rev  ..

 Align seq 1/1  to reverse of: AQ643973  from: 1  to: 127

       1 AlaLysAlaProGlyAlaSer 7
         |||||||||||||||||||||
      44 GCTAAAGCACCAGGTGCATCA 24

seq_name: gb_est46:AW485944

seq_documentation_block:
LOCUS        AW485944       132 bp    mRNA             EST       04-FEB-2000
DEFINITION  CM0.L10069.gi81299.144.a04.LT0069 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW485944
VERSION     AW485944.1  GI:6890604
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 132)
  AUTHORS   Gish,W. (http://www.tedwig.wustl.edu/ORESTES.
```

```
  TITLE     The FAPESP/LICR Human Cancer Genome Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509 cep, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/cgi-bin/db/dbtmp?p1=2&p2=2&p3=
            301299-144-a04&t3=1999-12-30&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 18
            High quality sequence stop: 68.
FEATURES             Location/Qualifiers
     source          1..132
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib "LIB069"
                     /dev_stage "Adult"
                     /note="Organ: leiomios; Vector: pmcRB; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       44 a      24 c      25 g      39 t
ORIGIN


alignment_scores:
            Quality:     7.00             Length:       7
              Ratio:    1.000              Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_75_83 x AW385944    ..

 Align seq 1/1  to: AW385944  from: 1  to: 132

        2 LysAlaProGlyAlaSerSer 8
          ||||||||||||||||||||||
        8 AAGGCTCCTGGTGCATCTTCC 28

seq_name: qb_est42:AW062078

seq_documentation_block:
LOCUS       AW062078       135 bp    mRNA            EST       05-OCT-1999
DEFINITION  660009D12.y1 660   Mixed stages of anther and pollen Zea mays cDNA,
            mRNA sequence.
ACCESSION   AW062078
VERSION     AW062078.1  GI:6012641
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 135)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave  Palo Alto  CA 94304  USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 660009  row: D  column: 12.
FEATURES             Location/Qualifiers
     source          1..135
                     /organism "Zea mays"
                     /cultivar "Ohio43"
                     /db_xref "taxon:4577"
                     /clone_lib "660 - Mixed stages of anther and pollen"
                     /tissue_type "whole premeiotic anthers to pollen shed"
                     /dev_stage "premeiotic anthers to pollen shed"
                     /lab_host "XL1-BLUE"
                     /note "Organ: anthers; Vector: Lambda Zap; Site_1: EcoRI;
                     Site_2: XhoI; Anther and pollen cDNA library.
                     Directionally sequenced with 5' end at the EcoRI site.
                     Created by Amie Franklin."
BASE COUNT       19 a      58 c      35 g      23 t
ORIGIN


alignment_scores:
            Quality:     7.00             Length:       7
              Ratio:    1.000              Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_75_83 x AW062078    ..

 Align seq 1/1  to: AW062078  from: 1  to: 135

        8 AlaProGlyAlaSerSerArg 9
          |||||||||||||||||||||
       77 GGCCCTGGCGCGAGTTCTCGC 97

seq_name: qb_est47:AW434511

seq_documentation_block:
LOCUS       AW434511       138 bp    mRNA            EST       11-FEB-2000
DEFINITION  UI-R-BJ0p-aff-a-02-0-UI.s1 UI-R-BJ0p-Rattus norvegicus cDNA clone
            UI-R-BJ0p-aff-a-02-0-UI 3', mRNA sequence.
ACCESSION   AW434511
VERSION     AW434511.1  GI:6965818
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 138)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            cDNA-dl track not found, Not 1 site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.B.
            Soares Lab clone distribution: clones will be available through
            Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1..138
                     /organism "Rattus norvegicus"
                     /strain "Sprague-Dawley"
                     /db_xref "taxon:10116"
                     /clone "UI-R-BJ0p-aff-a-02-0-UI"
                     /clone_lib "UI-R-BJ0p"
```

```
                     /dev_stage "adult"
                     /lab_host "DH10B (Life Technologies)"
                     /note "Vector: pT7T3D Pac (Pharmacia) with a modified
                     polylinker, Site 1: Not I, Site 2: Eco RI, The UI R B3Op
                     library is a subtracted library derived from the UI R AA1,
                     UI R-AB1, UI R-AC1, UI R AD1, UI-R-AE1, UI-R AF1, and
                     UI R-AG1 libraries. These libraries represent tissues from
                     rat atrium at 16.5 dpc, ventricle at 16.5 dpc, AV canal
                     at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV
                     canal at 15 dpc, and ventricle at 13 dpc. The tag is a
                     string of 5 6 nucleotides present between the Not I site
                     and the oligo dT track. The library was constructed as
                     described by Bonaldo, Lennon and Soares, Genome Research
                     6: 791 806, 1996.
                     TAG_LIB UI R B3Op
                     TAG_TISSUE AV canal at 16.5 dpc
                     TAG_SEQ GAAcC"
BASE COUNT        21 a      61 c      44 g      13 t
ORIGIN


alignment scores:
             Quality:     7.00              Length:       7
               Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2 copy 75_84 x AW434511/rev    ..

 Align seq 1/1  to reverse of: AW434511  from: 1  to: 138

        8 AlaProGlyAlaSerSerArg 9
          |||||||||||||||||||||
       47 GCGCCCGGGCCCTCGTCGCGA 27

seq name: qb qss84:AZ918723

seq documentation block:
LOCUS        AZ918723       139 bp    DNA              GSS        20-MAR-2001
DEFINITION   1006006D11.2EL_x1 1006   RescueMu Grid G Zea mays genomic, DNA
             sequence.
ACCESSION    AZ918723
VERSION      AZ918723.1  GI:13388007
KEYWORDS     GSS.
SOURCE       Zea mays.
  ORGANISM   Zea mays
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
             clade; Panicoideae; Andropogoneae; Zea.
REFERENCE    1  (bases 1 to 139)
  AUTHORS    Walbot,V.
  TITLE      Maize genomic sequences found using engineered RescueMu transposon
  JOURNAL    Unpublished (2001)
COMMENT      Contact: Walbot V
             Department of Biological Sciences
             Stanford University
             855 California Ave, Palo Alto, CA 94304, USA
             Tel: 650 723 2227
             Fax: 650 725 8221
             Email: walbot@stanford.edu
             Possible ligation site of ends cut by 2 different endonucleases.
             Reverse complemented post ligation sequence from source sequence.
             Plate: 1006006  row: D  column: 11
             Class: transposon tagged.
FEATURES             Location/Qualifiers
     source          1..139
                     /organism "Zea mays"
                     /cultivar "mix 1 ta 5 pr  x 5 W22/A554/B73"
                     /db_xref "taxon:4577"
                     /clone_lib "1006   RescueMu Grid G"
                     /tissue_type "leaf"
                     /dev_stage "adult"
                     /lab_host "DH10B"
                     /note "organ: leaf; Vector: RescueMu (engineered from
                     pBluescript backbone); Site 1: BamHI; Site 2: BglII;
                     RescueMu is a 4.9 kb, modified maize Mu transposon
                     designed to allow plasmid rescue from total genomic DNA.
                     Mu elements insert preferentially into transcription
                     units. For more information on RescueMu, go to the web
                     site 'www.zmdb.iastate.edu' and follow the links for
                     'RescueMu.' Grid G was grown at Stanford in 2000. DNA was
                     extracted from leaf punches, double digested using BamHI
                     and BglII, and ligated to form circular plasmids. DH10B
                     cells were transformed and then screened on LB plates with
                     ampicillin."
BASE COUNT        14 a      48 c      61 g      16 t
ORIGIN


alignment scores:
             Quality:     7.00              Length:       7
               Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2_copy 75_83 x AZ918723/rev    ..

 Align seq 1/1  to reverse of: AZ918723  from: 1  to: 139

        8 AlaProGlyAlaSerSerArg 9
          |||| ||||||||||||||||
      146 GCGCCCGGGCGCAGCTCCCGC 115
```

```
oM of: US-09-126-945B-2_COPY_75_83 to: GenEmbl:*    out_format : pts

Date: Sep 14, 2001  9:46 PM

About: Results were produced by the GenCore software, version 4.5.
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
  MODEL frame=_p2n.model  DEV xlp
  -Q "/cgi-bin/SPT_sp_cgi/%S96%26%45/tmp/sc 28***** 1 14_51 1*667/app_query.fasta_1.1267
  -DB GenEmbl -QFMT fastap -SUFFIX olip2n.rge  GAPOP 4.500
  -GAPEXT 0.050 -MINMATCH 0.100 -LOOPCUT 0.000  LOOPEXT 0.000
  QGAPOP 4.500 -QGAPEXT 0.050 -XGAPOP=60.000 -XGAPEXT=60.000
  -FGAPOP 6.000 -FGAPEXT=7.000 -YGAPOP 60.000 -YGAPEXT=60.000
  -DELOP 6.000 -DELEXT 7.000 -START 1 -MATRIX oligo
  -TRANS human40.cdi -LIST=45 -DOCALIGN 200 -THR_SCORE quality
  -THR_MIN 1 -ALIGN 15 -MODE LOCAL -QUIET pts -NORM qc  MINLEN 9
  -MAXLEN 20000000000 -USER=US09126945_*CGN1_1_7504 -NCPU 6
  -ICPU 3 -LONGLOG -NO_XLPXY -WAIT -THREADS 1

Search information block:
Query: US-09-126-945B-2_COPY_75_83
Query length: 9
Database: GenEmbl:*
Database sequences: 1344157
Database length: 856060004
Search time (sec): 6504.700000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000


score_list:
Sequence              Strd Orig    ZScore    EScore Len  ! Documentation ..
qb_rol:MUSITS          +     9.00  152.18   3.0e-07   218  ! M15652 Mouse (B86) insertion
qb_pr9:HSDNASAAF       -     9.00  151.08   8.5e-07   279  ! Z22243 H.sapiens DNA sequence,
qb_rol:MMWNT10B1       +     9.00  146.85   6.0e-06   726  ! U30462 Mus musculus wnt-10B gen
qb_pat1:AX048148       +     9.00  142.60   1.6e-05  1894  ! AX048148 Sequence 1 from Paten
qb_pr4:AF071538        +     9.00  142.60   1.6e-05  1894  ! AF071538 Homo sapiens Ets tran
qb_pr1:AB031549        +     9.00  142.60   1.6e-05  1898  ! AB031549 Homo sapiens PSE mRNA
qb_rol:MMU30464        +     9.00  141.91   1.1e-04  2215  ! U30464 Mus musculus wnt-10B mR
qb_rol:AF029307        +     9.00  140.50   1.4e-04  3848  ! AF029307 Mus musculus mammary
qb_pat1:AX048159       +     9.00  140.12   1.4e-04  3817  ! AX048159 Sequence 12 from Pate
qb_vil:AF033807        +     9.00  135.80   2.5e-04  8805  ! AF033807 Mouse mammary tumor v
qb_vi2:MMIPROCG        +     9.00  135.18   2.7e-04 10125  ! M15122 Mouse mammary tumor vi
qb_htq16:AC074221      -     9.00  124.32   1.0e-03 117655  ! AC074221 Mus musculus chromo
qb_pr6:AL157472        -     9.00  123.23   1.2e-03 150660  ! AL157472 Human DNA sequence
qb_bal:AE005854        +     8.00  118.65   2.2e-03 11094  ! AE005854 Caulobacter crescent
qb_htq10:AC025796      -     8.00  106.40   1.1e-03 176612  ! AC025796 Homo sapiens clone
qb_htq6:AC018863       -     8.00  106.17   1.1e-03 186053  ! AC018863 Homo sapiens chromo
qb_htq4:AC013799       -     8.00  106.06   1.1e-03 190650  ! AC013799 Homo sapiens clone
qb_htq13:AC053603      +     7.00  114.10   404.82   810  ! AC053603 Giardia intestinalis
qb_om:AF091609         +     7.00  113.97   411.87   885  ! AF091609 Sus scrofa clone pkgs1
qb_htq18:AC089132      -     7.00  113.71   425.97   886  ! AC089132 Giardia intestinalis
qb_htq13:AC053105      +     7.00  113.34   446.34   962  ! AC053105 Giardia intestinalis
qb_sts1:CNS06XKY       +     7.00  113.27   450.28   977  ! AL419864 T7 end of clone AY0AA0
qb_htq12:AC028752      -     7.00  113.17   456.27  1030  ! AC028752 Giardia intestinalis
qb_pl1:AE130973        +     7.00  112.42   502.20  1184  ! AE130973 Arabidopsis thaliana
qb_pat2:E05171         +     7.00  112.04   527.71  1292  ! E05171 cDNA to protein coding s
qb_pr10:HUMNFIL6BA     +     7.00  112.03   527.95  1293  ! M28667 Human NF-IL6 beta prote
qb_ba2:ECOSECE         -     7.00  111.75   547.89  1380  ! M38010 E.coli secE-nusG operon
qb_inl:AB010263        -     7.00  111.66   553.89  1407  ! AB010263 Drosophila melanogast
qb_pr10:S63168         +     7.00  111.11   594.56  1594  ! S63168 CCAAT/enhancer-binding
qb_ba3:MXTAH           -     7.00  110.87   612.78  1681  ! AJ242955 Myxococcus xanthus cy
qb_ba3:PSVSRA          -     7.00  110.76   621.83  1725  ! U02041 Pseudomonas solanacearu
qb_pr4:AF052130        +     7.00  110.54   639.05  1810  ! AF052130 Homo sapiens clone 23
qb_pat2:AX073574       -     7.00  110.13   673.62  1986  ! AX073574 Sequence 26 from Pate
qb_ba3:SAUNSTDNA       -     7.00  109.80   703.18  2142  ! X70975 S.ambofaciens DNA seque
qb_ba3:RSU87369        +     7.00  108.54   818.20  2946  ! U87369 Rhodobacter sphaeroides
qb_pr7:BC003617        +     7.00  108.19   864.06  3079  ! BC003617 Homo sapiens, RAB6, m
qb_rol:BC003210        +     7.00  108.15   868.67  3108  ! BC003210 Mus musculus, hypothe
qb_ro2:RATMAISTOM      -     7.00  107.12   991.76  3925  ! L11236 Rattus rattus mRNA sequ
qb_pat2:AX081469       +     7.00  106.38   1.1e-03  4632  ! AX081469 Sequence 11 from Pate
qb_pr5:AF250860        -     7.00  106.31   1.1e-03  4713  ! AF250860 Bos taurus actin
qb_ba3:PSEY18C27       -     7.00  105.39   1.2e-03  5795  ! Y18C27 Pseudomonas sp. pobA
qb_rol:MMU24579        +     7.00  104.75   1.3e-03  6703  ! AJ245739 Mus musculus parti
qb_om:OAIRTB           -     7.00  103.28   1.6e-03  9444  ! Z18061 Ovalbumin-like trichohyalin
qb_om:RABGL-B.N        -     7.00  102.71   1.7e-03 10621  ! M7414, Rabbit beta 1 globi
qb_bal:AE005969        -     7.00  102.56   1.8e-03 10990  ! AE005969 Caulobacter cresc

seq_name: qb_rol:MUSITS

seq_documentation_block:
LOCUS       MUSITS          218 bp    DNA             ROD       27-APR-1993
DEFINITION  Mouse (B86) insertion target site for MMTV retrovirus.
ACCESSION   M15652
VERSION     M15652.1  GI:198464
KEYWORDS    insertion site; retroviral insertion site.
SOURCE      Mouse (B86) mammary tumor DNA.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 218)
  AUTHORS   Moore,R., Dixon,M., Smith,R., Peters,G. and Dickson,C.
  TITLE     Complete nucleotide sequence of a milk transmitted mouse mammary
            tumor virus: two frameshift suppression events are required for
            translation of gag and pol
  JOURNAL   J. Virol. 61, 480-490 (1987)
  MEDLINE   87212944
COMMENT     Draft entry and computer readable copy of sequence in [1] were
            kindly provided by C.Dickson, 05-MAY-1987.
            This sequence was reconstructed from figure 1 in [1] by deleting
            the proviral DNA and one of the copies of the 6 bp target
            duplication.
FEATURES             Location/Qualifiers
     source          1..218
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     misc_feature    120..125
                     /note="target site (duplicated on integration)"
BASE COUNT       27 a     56 c     88 g     47 t
ORIGIN      Unreported.


alignment scores:
            Quality:    9.00              Length:      9
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_75_83 x MUSITS  ..

 Align seq 1/1  to: MUSITS  from: 1  to: 218

       1 AlaLysAlaProGlyAlaSerSerArg 9
         !!!!!!!!!!!!!!!!!!!!!!!!!!!
     162 GCCAAAGCCCCCGGTAGCTTCGTCGCGG 188

seq_name: qb_pr9:HSDNASAAF

seq_documentation_block:
LOCUS       HSDNASAAF     279 bp    DNA             PRI       01-JAN-1995
DEFINITION  H.sapiens DNA sequence.
ACCESSION   Z22243
VERSION     Z22243.1  GI:608728
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 279)
  AUTHORS   Zabarovsky,E.R., Kashuba,V., Pettersson,P., Petrovic,N., Zaktaryev,V.,
            Grigoriev,P., Ibbotson,I., Bannikov,V., Erlandsson,R., Uhlen,M.,
            Kisselev,L. and Klein,G.
  TITLE     Shot-gun sequencing strategy for long range genome mapping: first
            results
```

```
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 279)
  AUTHORS   Zabarovsky,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (01 APR 1993) Zabarovsky E., Karolinska Institute, Tumor
            Biology, S 104 01 Stockholm, Sweden
FEATURES             Location/Qualifiers
     source          1..279
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="chromosome 3"
                     /germline
                     /clone="924Z025R"
                     /cell_line="MCH903.1"
                     /clone_lib="924.4"
BASE COUNT       67 a    106 c     74 g     28 t      4 others
ORIGIN


alignment scores:
            Quality:    9.00              Length:      9
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_75_83 x BSDNASAAF/rc  ..

 Align seq 1/1  to reverse of: BSDNASAAF  from: 1  to: 279

       1 AlaLysAlaProGlyAlaSerSerArg 9
         ||||||||||||||||||||||||||
      94 GCCAAAGCCCCCGGTGCTTCCGTCGCGG 68

Seq name: qb_rol:MMWNT10B1

Seq documentation block:
LOCUS       MMWNT10B1      726 bp    DNA             ROD       15-OCT-1996
DEFINITION  Mus musculus wnt-10B gene, 5' region and exon 1.
ACCESSION   U30462
VERSION     U30462.1  GI:1134445
KEYWORDS    .
SEGMENT     1 of 2
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Moore,R., Dixon,M., Smith,R., Peters,G. and Dickson,C.
  TITLE     Complete nucleotide sequence of a milk-transmitted mouse mammary
            tumor virus: two frameshift suppression events are required for
            translation of gag and pol
  JOURNAL   J. Virol. 61 (2), 480-490 (1987)
  MEDLINE   87112944
REFERENCE   2  (bases 1 to 726)
  AUTHORS   Hardiman,G., Albright,S., Tsunoda,J., McClanahan,T. and Lee,F.
  TITLE     The mouse Wnt-10B gene isolated from helper T cells is widely
            expressed and a possible oncogene in BR6 mouse mammary
            tumorigenesis
  JOURNAL   Gene 172 (2), 199-205 (1996)
  MEDLINE   96269404
REFERENCE   3  (bases 1 to 726)
  AUTHORS   Hardiman,G., Albright,S., Tsunoda,J., McClanahan,T. and Lee,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUN-1995) Gary Hardiman, Molecular Biology, DNAX, 901
            California Ave., Palo Alto, CA 94304-1104, USA
FEATURES             Location/Qualifiers
     source          1..726
                     /organism="Mus musculus"
                     /strain="BR6"
                     /db_xref="taxon:10090"
                     /tissue_type="BR6 mouse mammary tumor (designated W26)"
     exon            188..479
                     /gene="wnt-10B"
                     /number=1
     intron          480..>726
                     /gene="wnt-10B"
                     /number=1
BASE COUNT      169 a    289 c    217 g    131 t
ORIGIN


alignment scores:
            Quality:    9.00              Length:      9
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_75_83 x MMWNT10B1  ..

 Align seq 1/1  to: MMWNT10B1  from: 1  to: 726

       1 AlaLysAlaProGlyAlaSerSerArg 9
         ||||||||||||||||||||||||||
     224 GCCAAAGCCCCCGGTGCTTCCGTCGCGG 250

Seq name: qb_pat1:AX048148

Seq documentation block:
LOCUS       AX048148      1894 bp    DNA             PAT       15-DEC-2000
DEFINITION  Sequence 1 from Patent WO0070092.
ACCESSION   AX048148
VERSION     AX048148.1  GI:11876967
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1894)
  AUTHORS   Kaufmann,J. and Xin,H.
  TITLE     Expression of ets domain proteins in cancer
  JOURNAL   Patent: WO 0070092-A 1 23-NOV-2000;
            Chiron Corporation (US)
FEATURES             Location/Qualifiers
     source          1..1894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             416..1423
                     /note="Human JKETS."
                     /codon_start=1
                     /protein_id="CAC19002.1"
                     /db_xref="GI:11876968"
                     /translation="MGSASPGLSSVSPSHLLIPPPQVSPQTGLEKAAAGAVGLERKDWS
                     PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAPGASSREFPPEEFEQCPVTDSQAPA
                     GSLDLVPGGLTLEEHSLEQVQSMVVGEVELKDTETACKLLNITADPMDWSPSNVQKWLL
                     WTFIQQYRLIPPMGKAFQELAGKELCAMSEEQFPQRSPLQGPVLHAHETDWKSAAWMEK
                     ISPCATHYCASTSEESWIDSEVDSSCSGQQCTHLWQFLKKELLKPHSYGRETRWLNKEK
                     CIRKELRSAQVAELWGIRKNKPAMNYDKLSRSLRQYYEKGIIRKPTDISQRLVYQFVHP
                     I"
BASE COUNT      568 a    658 c    571 g    302 t
ORIGIN


alignment scores:
            Quality:    9.00              Length:      9
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_75_83 x AX048148  ..

 Align seq 1/1  to: AX048148  from: 1  to: 1894

       1 AlaLysAlaProGlyAlaSerSerArg 9
         ||||||||||||||||||||||||||
     636 GCCAAAGCCCCCGGGGCCAGCAGTCGC 664
```

```
seq_name: qb_pr4:AF071538

seq_documentation_block:
LOCUS       AF071538     1894 bp    mRNA            PRI       15-JAN-2000
DEFINITION  Homo sapiens Ets transcription factor PDEF (PDEF) mRNA, complete
            cds.
ACCESSION   AF071538
VERSION     AF071538.1  GI:4007417
KEYWORDS    .
SOURCE      human
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1894)
  AUTHORS   Oettgen,P., Finger,E., Sun,Z., Akbarali,Y., Thamrongsak,U.,
            Boltax,J., Grall,F., Dube,A., Weiss,A., Brown,L., Quinn,G., Kas,E.,
            Endress,G., Kunsch,C. and Libermann,T.A.
  TITLE     PDEF, a novel prostate epithelium specific Ets transcription
            factor, interacts with the androgen receptor and activates
            prostate-specific antigen gene expression
  JOURNAL   J. Biol. Chem. 275 (2), 1216-1225 (2000)
  MEDLINE   20092894
REFERENCE   2  (bases 1 to 1894)
  AUTHORS   Libermann,T.A., Oettgen,P., Kunsch,C. and Finger,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUN-1998) Medicine, Beth Israel Deaconess Medical
            Center, 330 Brookline Ave., Boston, MA 02215, USA
FEATURES             Location/Qualifiers
     source          1..1894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6p21.3"
                     /tissue_type="prostate"
     gene            1..1894
                     /gene="PDEF"
     CDS             416..1423
                     /gene="PDEF"
                     /note="prostate epithelium-specific"
                     /codon_start=1
                     /product="Ets transcription factor PDEF"
                     /protein_id="AAC95296.1"
                     /db_xref="GI:4007418"
                     /translation="MASASGLSAVSLSGLFLEQVSREGEKAAAGAVGLERRDWS
                     PSPPAIPEQGLSAFYLSYFDMLYPEDSSWAAKAPVASSKEEPPEEFEFQVTVDSQAFA
                     GSLDLVPGGLTLEEHSLEQVQSMVVGEVLKDIETACKLLNLIAEPMDWSPSNVQKWLL
                     WTEHQYRLEPMGKAFQELACPELACMSEEQEEQESPLGGQVLHAHEDLWFSAAWMKER
                     ISPGATHYCASTSPESWTDSFVDSSVSGQFHLWQFLLELLKPHSYGREFFRWLNEEF
                     GIFKIFDSAQVAPLWGIPKNRDAMNYDFLSRSLPQYPEQVAYPLVLVQFVHP
                     1"
BASE COUNT       368 a    653 c    571 g    302 t
ORIGIN


alignment_scores:
             Quality:     9.00              Length:       9
               Ratio:    1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_75_83 x AF071538  ..

 Align seq 1/1  to: AF071538  from: 1  to: 1894

        1 AlaLysAlaProGlyAlaSerSerArg 9
          |||||||||||||||||||||||||||
      638 GCCAAGGCCCCTGGGGCCAGCAGTCGG 664

seq_name: qb_pr1:AB031549

seq_documentation_block:
LOCUS       AB031549     1898 bp    mRNA            PRI       25-JAN-2000
DEFINITION  Homo sapiens PSE mRNA for prostate ets, complete cds.
ACCESSION   AB031549
VERSION     AB031549.1  GI:6721497
KEYWORDS    prostate ets.
SOURCE      Homo sapiens adult male prostate epithelium prostate cancer
            cell_line:PC-3 cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Yamada,N., Tamai,Y., Miyamoto,H. and Nozaki,M.
  TITLE     Cloning and expression of the mouse Pse gene encoding a prostate Ets
            family member
  JOURNAL   Gene 241 (2), 267-274 (2000)
  MEDLINE   20137504
REFERENCE   2  (bases 1 to 1898)
  AUTHORS   Nozaki,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2000) to the EMBL/GenBank/DDBJ databases. Masami
            Nozaki, Osaka University, Institute for Microbial Diseases; 3-1,
            Yamadaoka, Suita, Osaka 565-0871, Japan
            (E-mail:mnozaki@biken.osaka-u.ac.jp, Tel:81-6-879-8338,
            Fax:81-6-879-8339)
FEATURES             Location/Qualifiers
     source          1..1898
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="PC-3"
                     /cell_type="prostate cancer"
                     /chromosome="6"
                     /dev_stage="adult"
                     /map="6p21.2-21.3"
                     /sex="male"
                     /tissue_type="prostate epithelium"
                     /note="vector:pUC18"
     gene            406..1413
                     /gene="PSE"
     CDS             406..1413
                     /gene="PSE"
                     /codon_start=1
                     /product="prostate ets"
                     /protein_id="BAA89543.1"
                     /db_xref="GI:6721498"
                     /translation="MCSASPCLSSVSLSHLLLPPQVSREGEKAAAGAVGLERRDWS
                     [illegible]
                     GSLDLVPGGLTLEEHSLEQVQSMVVGEVLKDIETACKLLNLIAEPMDWSPSNVQKWLL
                     WTEHQYRLEPMGKAFQELACPELACMSEEQEEQESPLGGQVLHAHEDLWFSAAWMKER
                     ISPGATHYCASTSPESWTDSFVDSSVSGQFHLWQFLLELLKPHSYGREFFRWLNEEF
                     GIFKIFDSAQVAPLWGIPKNRDAMNYDFLSRSLPQYPEQVAYPLVLVQFVHP
                     1"
BASE COUNT       380 a    651 c    567 g    300 t
ORIGIN


alignment_scores:
             Quality:     9.00              Length:       9
               Ratio:    1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_75_83 x AB031549  ..

 Align seq 1/1  to: AB031549  from: 1  to: 1898

        1 AlaLysAlaProGlyAlaSerSerArg 9
          |||||||||||||||||||||||||||
      628 GCCAAGGCCCCTGGGGCCAGCAGTCGG 654

seq_name: qb_rol:MM030464

seq_documentation_block:
LOCUS       MM030464     2215 bp    mRNA            ROD       28-DEC-1995
DEFINITION  Mus musculus wnt 10B mRNA, complete cds.
ACCESSION   030464
```

```
VERSION     U30464.1  GI:1136444
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorphi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2215)
  AUTHORS   Hardiman,G., Albright,S., Tsunoda,J., McClanahan,T and Lee,E.
  TITLE     Mouse wnt 10B isolated from helper T cells is widely expressed and
            a possible oncogene in B86 mouse mammary tumorigenesis
  JOURNAL   Unpublished (1995)
REFERENCE   2  (bases 1 to 2215)
  AUTHORS   Hardiman,G., Albright,S., Tsunoda,J., McClanahan,T and Lee,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (27 JUN 1995) Gary Hardiman, Molecular Biology, DNAX, 901
            California Ave, Palo Alto, CA 94304-1104, USA
FEATURES             Location/Qualifiers
     source          1..2215
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /cell_type="B?? T cell"
     gene            342..1501
                     /gene="wnt 10B"
     CDS             342..1501
                     /gene="wnt 10B"
                     /note="possible oncogene"
                     /codon_start=1
                     /protein_id="AAA84399.1"
                     /db_xref="GI:1136445"
                     /translation="MLETPPCFPFSCCALILFLALFSKAISNEILGKIWGRFLIA
                     NIVCLTLSGLSKRQLGLCERSPDVIASALQGLHIAVHECQHQLRDQRWNCSALEGGGR
                     LPHHSATLKRG"
     polyA_site      2215
                     /note="18 A nucleotides"
BASE COUNT      428 a    640 c    668 g    489 t
ORIGIN


alignment scores:
            Quality:     9.00              Length:       9
              Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2-COPY_75_83 x MM030464    ..

 Align seq 1/1  to: MM030464  from: 1  to: 2215

        1 AlaLysAlaProGlyAlaSerSerArg 9
          ||||||||||||||||||||||||||
       3/ GCCAAAGCCCCCGGTGCTTCCGTCGCGC 63

seq name: gb|rol:AF029407

seq documentation block:
LOCUS       AF029407      4048 bp    DNA             ROD       08-NOV-1997
DEFINITION  Mus musculus mammary gland growth regulator wnt10b (WNT10b) gene,
            partial cds.
ACCESSION   AF029407
VERSION     AF029407.1  GI:2598207
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorphi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 4048)
  AUTHORS   Lee,E.S., Lane,T.F., Kno,A., Shackleford,G.M. and Leder,P.
  TITLE     Insertional mutagenesis identifies a member of the Wnt gene family
            as a candidate oncogene in the mammary epithelium of int-2/Fgf-3
            transgenic mice
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 92 (6), 2268-2272 (1995)
  MEDLINE   95199333
REFERENCE   2  (bases 1 to 4048)
  AUTHORS   Lane,T.F. and Leder,P.
  TITLE     Wnt-10b directs hypermorphic development and transformation in
            mammary glands of male and female mice
  JOURNAL   Oncogene 15 (18), 2133-2144 (1997)
  MEDLINE   98054120
REFERENCE   3  (bases 1 to 4048)
  AUTHORS   Lane,T.F. and Leder,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-1997) Genetics, Harvard Medical School/DBMI, 200
            Longwood Ave, Boston, MA 02115, USA
FEATURES             Location/Qualifiers
     source          1..4048
                     /organism="Mus musculus"
                     /strain="129/Sv"
                     /db_xref="taxon:10090"
                     /chromosome="15"
                     /map="between Pmv-42 and D-ex 3"
     repeat_region   complement(491..541)
                     /rpt_type=tandem
                     /rpt_unit=CA
     mRNA            join(706..940,1791..1904,2014..>2278)
                     /gene="WNT10b"
                     /product="mammary gland growth regulator wnt10b"
     exon            706..940
                     /gene="WNT10b"
                     /note="exon 1a"
     gene            706..>2278
                     /gene="WNT10b"
     misc_feature    1387..1402
                     /gene="WNT10b"
                     /note="murine mammary virus integration site 'NX6.0'"
     mRNA            join(1533..1640,1791..1904,2014..>2278)
                     /gene="WNT10b"
                     /product="mammary gland growth regulator wnt10b"
     exon            1533..1640
                     /gene="WNT10b"
                     /note="exon 1b"
     misc_feature    1540..1555
                     /gene="WNT10b"
                     /note="murine mammary virus insertion site 'NX/Int-2'"
     exon            1791..1904
                     /gene="WNT10b"
                     /number=2
     CDS             join(1841..1904,2014..>2278)
                     /gene="WNT10b"
                     /note="wnt1/wg homolog, secreted growth factor"
                     /codon_start=1
                     /product="mammary gland growth regulator wnt10b"
                     /protein_id="AAC53429.1"
                     /db_xref="GI:2598208"
                     /translation="MLEEPRSRFPFLCLAGILFLALFSKAISNEILGKIWGRFLIA
                     NIVCLTLSGLSKRQLGLCERSPDVIASALQGLHIAVHECQHQLRDQRWNCSALEGGGR
                     LPHHSATLKRG"
     exon            2014..>2278
                     /gene="WNT10b"
                     /number=3
BASE COUNT      585 a    871 c   1012 g    582 t
ORIGIN


alignment scores:
            Quality:     9.00              Length:       9
              Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2-COPY_75_83 x AF029407    ..

 Align seq 1/1  to: AF029407  from: 1  to: 4048
```

```
        1 AlaLysAlaProGlyAlaSerSerArg 9
          |||||||||||||||||||||||||||
      684 GCCAAAGCCCCCGGTGCTTCGTCCCGG 710

seq_name: qb_pat1:AX048159

seq_documentation_block:
LOCUS       AX048159      3417 bp    DNA             PAT       15-DEC-2000
DEFINITION  Sequence 12 from Patent WO0070092.
ACCESSION   AX048159
VERSION     AX048159.1  GI:11876976
KEYWORDS
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3417)
  AUTHORS   Kaufmann,J. and Xin,H.
  TITLE     Expression of ets-domain proteins in cancer
  JOURNAL   Patent: WO 0070092 A 12-22-NOV-2000
            Chiron Corporation (US)
FEATURES             Location/Qualifiers
     source          1..3417
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="GSEP modified promoter with gag gene 5' UTR, SA, SD"
     promoter        1..1380
                     /note="Promoter"
     CDS             1815..2822
                     /note="GSEP coding sequence"
                     /codon_start=1
                     /protein_id="AAC82556.1"
                     /db_xref="GI:11876976"
                     /translation="MCVSCRQGFEVSVEPCLEVVQQALSSPETPRPFQAAAQPAVGLEEPPWS
                     PSPPATPEQGLSAFYLSYFDMLYPEDCGWAAKAFGASSKEEPEEPEQCPVTPSQAPA
                     GSLDLVPGGLTLEEHSLEQVQSMVVGPVEKGTFTACKLNSAADMEWSSSNVSKWTT
                     WTFHQYRLPPMQKAPQFTAGKFLGAMSFPQVPPQQPDGVEHARLTLWFSAAWMPEP
                     TSPGAIHYCASTSEESWTDSEVDSSCSGQPTHLWQFLKPILLKPHSYCPEYRWTNKEK
                     GIFKIPDSAQVARTWCIRKNKPAMNYDKLSRSLRYYYKKCTKSSPDLSQKLVYQPVHP
                     I"
BASE COUNT      710 a    1026 c     970 g     611 t
ORIGIN


alignment_scores:
            Quality:     9.00            Length:       9
              Ratio:    1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_75_83 x AX048159  ..

 Align_seq 1/1  to: AX048159  from: 1  to: 3417

        1 AlaLysAlaProGlyAlaSerSerArg 9
          |||||||||||||||||||||||||||
    2037 GCCAAGCCCCCTGGGGCCAGCAGTCGG 2063

seq_name: qb_vil:AF033807

seq_documentation_block:
LOCUS       AF033807      8805 bp    RNA             VRL       03-DEC-1998
DEFINITION  Mouse mammary tumor virus complete proviral genome.
ACCESSION   AF033807
VERSION     AF033807.1  GI:2801453
KEYWORDS    .
SOURCE      Mouse mammary tumor virus.
  ORGANISM  Mouse mammary tumor virus
            Viruses; Retroid viruses; Retroviridae; Betaretrovirus; Mammalian
            type B retroviruses.
REFERENCE   1  (bases 1 to 8805)
  AUTHORS   Petropoulos,C.J.
  TITLE     Appendix 2: Retroviral taxonomy, protein structure, sequences, and
            genetic maps
  JOURNAL   (in) Coffin,J.M. (Ed.);
            RETROVIRUSES; 757;
            Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
            NY, USA (1997)
REFERENCE   2  (bases 1 to 8805)
  AUTHORS   Chappey,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-NOV-1997) NIH, NLM, Rockville Pike, Bethesda, MD
            20894, USA
FEATURES             Location/Qualifiers
     source          1..8805
                     /organism="Mouse mammary tumor virus"
                     /db_xref="taxon:11757"
     mRNA            1..8805
                     /gene="gag-pro-pol"
     5'UTR           16..153
     gene            313..2088
                     /gene="gag"
     CDS             join(313..2082,2082..2891,2891..5578)
                     /gene="gag-pro-pol"
                     /codon_start=1
                     /product="Pr160"
                     /protein_id="AAC82556.1"
                     /db_xref="GI:2801455"
                     /translation="MCVSCRQGFEVSVEPCLEVVQQALSSPETPRPFQAAAQPAVGLEEPPWS
                     WFPEEGGLNLQDWKRVGREMKRYAAEHCTDSTPKQAYFTWLQLREILLEQSDLVLLSA
                     EAKSVTEEELEEGLTGLLSTSSQEKTYGTRGTAYAETDTEVDKLSEHTYDEPYFEKEK
                     AQPNEIEHLHVEELFEVVLPKELNSSGREKLKSFAFLATLHWNDDDLSPELWDDLEEQAA
                     RVPIPPDVVPKEFVVKEEPVSALPERPLEPPVGEACAMALAPEKQDLIETFPVVFMGE
                     SPSPDVVIWRTEYEEKSFEMVQKAAGRREQFVSLIDMIFCSQPELSPSSQPESPSVPA
                     DVLTNAVLAWKAIPPPGVKKIVLAGLKQGNEESYEIFTSKLEEAVYRMMPKDGQSDIL
                     IKQLAWENANSLCQDLIRPIRKTGTIQDYIRACLIDASPAVVQGMAYAAAMKGQKYSIF
                     VKQTYGGQKGQGQAEGPVCFSCGKIGHIPKDCKDKGSKEAAPPGLCPKCKKGYHWKSE
                     CNSKPNDGNPLPTLEINAENSKNLVKGQSPSPAQKGDGVKGSGLNLEAPPFTIHDLP
                     KTFSGSAELLLSSPPGLCSSELVSVLEVLPKTSPLSLFGVNYEKHEVL
                     PSVPLPSPQVQELFVMVKAAKNAVIHKGGPIAQLLLLPYLKLPNPVIKEERGSEQFGS
                     ISHVHWVQELISDSKPMLHIYLNGRRFLGLLDTGADKTCIAGRDWPANWPIHQIESSLQ
                     GLGMACGVAKSSQPIPWQHFDKSGTIHPFVIPTLPETIWGRDIMKDIKVRLMIDSPDD
                     SQDLMIGAIESNLEFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESN
                     SPWNTPVFVIKKKSGKWRLLQDLRAVNATMHDMGALQPGLPSPVAVPKGWETIIIDLQ
                     DCFFNIKLHPEDCKKFAFSVPSPNFKEPYQRFQWKVLPQGMKNSPTICQLKEVDKAIT
                     VRDKYQDSYIVHYMDDILLAHPSKSTVDELITSMIQALNKHGLVVSIEKTQKYDNLKY
                     LGTHIQGDSVSYQKLQIKLDKLKLLNDFQKLIGNINWIKPFLKLTIGELKPLFEETLNG
                     DSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSICTLKTEYTPTACLWQDG
                     VVFWIPLPHTSPKVITPYDTFCTQLIIKCPHPSKPFLFSKDPDYIVVPVYKVQPDLLLQ
                     EKEDWPLSLIGLGEVFPHLPKDPLLIFTLQTATIFPPHMISTIPLEKGIVIFIDGSAN
                     GRSVTYIQGREPTIKENIQNTAQQAEIVAVIIAFEEVSQPPNLYIDSKYVIGLFPEEP
                     IAILSPKIKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLIGPLAQGNAYADSLIRILI
                     ALPSAQPSHAIHHQNAAALRPQFHTIPFQAPFIVKTCPNCHWGHAPLQCVNPPGLEP
                     RVLWQMDVTHVSEFGKLKYVHVIVDTYSHFIFAIAKTGEATKDVLQHLAQSFAYMGIP
                     QKIKIDNAPAYVSKSIQEFLAKWEISHVTGIPYNPQSQAIVERTHQNIKAQLNKLQEA
                     GKYYTPDHLIAHALFVLNMNMNQGHIAAERPHWGPTSAQFFFMVMWWLILGSWFGP
                     DVLITAGRGYACVFPQDAETPIWVPDRFIKPETEKEFATFTPGIAPKIPPPEQPQQE
                     SPKNESSPHQREDGLATSAGVDLRSGGQP"
     CDS             join(313..2082,2082..2894)
                     /gene="gag-pro"
                     /codon_start=1
                     /product="Pr110"
                     /protein_id="AAC82555.1"
                     /db_xref="GI:2801454"
                     /translation="MCVSCRQGFEVSVEPCLEVVQQALSSPETPRPFQAAAQPAVGLEEPPWS
                     WFPEEGGLNLQDWKRVGREMKRYAAEHCTDSTPKQAYFTWLQLREILLEQSDLVLLSA
                     EAKSVTEEELEEGLTGLLSTSSQEKTYGTRGTAYAETDTEVDKLSEHTYDEPYFEKEK
                     AQPNEIEHLHVEELFEVVLPKELNSSGREKLKSFAFLATLHWNDDDLSPELWDDLEEQAA
                     RVPIPPDVVPKEFVVKEEPVSALPERPLEPPVGEACAMALAPEKQDLIETFPVVFMGE
                     SPSPDVVIWRTEYEEKSFEMVQKAAGRREQFVSLIDMIFCSQPELSPSSQPESPSVPA
                     DVLTNAVLAWKAIPPPGVKKIVLAGLKQGNEESYEIFTSKLEEAVYRMMPKDGQSDIL
                     IKQLAWENANSLCQDLIRPIRKTGTIQDYIRACLIDASPAVVQGMAYAAAMKGQKYSIF
                     VKQTYGGQKGQGQAEGPVCFSCGKIGHIPKDCKDKGSKEAAPPGLCPKCKKGYHWKSE
```

```
                     CKSFEFEKSCNPLLPFLETNAENSKNL"
     CDS             413..2088
                     /gene "gag"
                     /codon_start=1
                     /product="Pr77"
                     /protein_id="AAC82557.1"
                     /db_xref="GI:2801456"
                     /translation="MGVSGSKQGPTFEVSVEGPLTSSPSTHVFKSAIFEYGAKRWTR
                     WFPEEGGLNLQGWFPVCPEMPPYAAEHCHLSTEPQAYFPLWLQLRPLFGNSSLVLTSA
                     EAKSVTEEELRECGLTGLLSTSSQEKTYGTRCTAYAEIDIEVDKLSEHIYDEPYEEKEK
                     ADKNEEKDHVKPLFERVVQPFFNSEGPEPFEFTSFAFLATIWNEFRQLSPEDWSQLEEQAA
                     HYHDDDELTLPVKRKVVEKRKPQALPRKPLPFPVGEAGAMAEAPEEGGLLETEPVVFMGE
                     SDEDDTPVWPPLPELKLTKEEQSAVKEMGDSAPYTLQVVDMVASQWETDSDWHQIAKAI
                     LSPGDYYVLWPTEYFEPKSKEMVQKAAGKRKVSSLMELGTGQPELSPSSQTKLSKDVLK
                     DVTTNAVLAWKAIPPPGVKKLTVLAGLKQGNREESYLTPLSKLEEAVYKMMFQGCGSPTL
                     IKQLAWENANSLSQDLIPPTREPTGLTQDYTPAQLDASPAVVQGMAYAAAMPQGKYSTF
                     VKQTYQGGKEQQGAKGPVPESCQKLQHEMKQVFEEGSKRPAPPQTGPPQKEQSGWRSR
                     CKSKEDKDGNPLPPLETNAENSKNL"
     mat_peptide     416..609
                     /gene "gag"
                     /product="p10 MA"
     mat_peptide     610..897
                     /gene "gag"
                     /product="pp21 XX"
     mat_peptide     898..996
                     /gene "gag"
                     /product="p3"
     mat_peptide     997..1071
                     /gene "gag"
                     /product="p8"
     mat_peptide     1069..1800
                     /gene "gag"
                     /product="p27 CA"
     mat_peptide     join(1801..2082,2082..2546)
                     /product="p14 NC"
     gene            2082..2894
                     /gene "pro"
     mat_peptide     join(2544..2891,2891..5575)
                     /product="p40DU-p13PR-RT-IN"
     gene            2891..5579
                     /gene "pol"
     mRNA            5364..7430
                     /gene "env"
                     /product="gPr73"
     gene            5364..7430
                     /gene "env"
     CDS             5364..7430
                     /gene "env"
                     /codon_start=1
                     /product="gPr73"
                     /protein_id="AAC82558.1"
                     /db_xref="GI:2801457"
                     /translation="MPPRQPSSPPPSSPLSLEESPLQPQPHALPPPPPPVPGESYW
                     RMNLVPIKEKFAWQHLQALTSEAEEVLKTSQTPQTSLTLELALESVECPPPVPGESYW
                     AYLPKPPTLHPVGWGSTDPIRVLTNQTMYLGGSPDFHGFPRNMSGNVHFEGKSPFLFI
                     LSFSFVTTPQTPQVPKQVELRPGVEYKYNFNSLTIEEEQLQKANEFEV
                     PTEEKKLPFEYYGQQLAFKDAEWEIQLSAPPPWELPAPPQCVSESPKGALSLLLWDES
                     LPSPSVDQSDQIESKKNLFGNYTEPVNKEVHPWYEAGWVEPTWEWENSPEFPNPEDPT
                     ALVPHTELERLVAASRHLTLKRPGFQEHEMITPTSACVTYPYATLLGLFQLTDTEKRGS
                     TFHISCSSCRLINCLDSSAYDYAATLVKRPFYVLLPVDTGDEPWFDDSATQTFRYATD
                     LTRAFEVAATETSLALTALEPAVATAALVPPMQTAFEVNNCHENVEIALPPPFI
                     IDLKLEARLNALEEVVLEEGQDVANLFTEMETDCHANYDFTPVTELFYNALENWEEIE
                     ANLLGIWNDNETSYNIQELTNLISDMSKQHTDAVDLSGLAQSFANGVKALNPLHWTQY
                     FTETGVCALLLVTVLMTFPIVFQCLAKSLDQVLSSEINVLLLKKKKCGNAAFAALEMVT
                     PRVSYI"
     mat_peptide     5658..6785
                     /gene "env"
                     /product="gp52 SU"
     mat_peptide     6796..7427
                     /gene "env"
                     /product="gp46 TM"
     3'UTR           7427..8573
     mRNA            7478..8440
                     /gene "SAG"
                     /product="Pr48"
     gene            7478..8440
                     /gene "SAG"
     CDS             7478..8440
                     /gene "SAG"
                     /codon_start=1
                     /product="Pr48"
                     /protein_id="AAC82559.1"
                     /db_xref="GI:2801458"
                     /translation="MPRLQQKWLNSKEPTLLRGEAAKGLEETKEAQESQHERMSPSHKD
                     LLTLVPASTAALTGLIGEVAVKAEPALLLESNSSSVQGYNLNGGENSLELLKQGP
                     QPQSSSYKPQQPPSEIETRMLAKNYTEINKLNPTQKELVTMLRNESLESTLTQTQK
                     LEMGTENKKRHSTSVEEQVQGLEASGLEVKEGKKSALVKTGDKWWQPGTYRGPYIYRP
                     IDAPLPYTGKYILNFDKWVEVNGYKVLYKSLSLFEKEKLAKAPEPWEMETQLEEKNQMEQQ
                     VHDYIYLGTGMSSTWGKTFETTKEPIVAALIEHYSAKTYGMSYYD"
     BASE COUNT      2446 a  1825 c  1969 g  2415 t
     ORIGIN


     alignment scores:
                  Quality:     9.00              Length:       9
                    Ratio:  1.000                Gaps:       0
      Percent Similarity: 100.000   Percent Identity: 100.000

     alignment block:
      us-09-126-945b-2_copy_75_83 x AF033807  ..

      Align seq 1/.  to: AF033807  from: 1  to: 8805

             1 AlaLysAlaProGlyAlaSerSerArg 9
               |||| |||||||||||||||||||||
          8749 gGCAAAgcccCCGGTGCCTCCTCGTCGCGG 8775

     seq name: qb_vi2:MMTPROCG

     seq documentation block:
     LOCUS       MMTPROCG     10125 bp ss-RNA             VRL       02 AUG 1993
     DEFINITION  Mouse mammary tumor virus complete proviral genome and mouse DNA.
     ACCESSION   M15122
     VERSION     M15122.1  GI:332127
     KEYWORDS    complete genome; gag protein; long terminal repeat (LTR);
                 polymerase; protease.
     SOURCE      Mouse (BR6) mammary tumor proviral DNA.
       ORGANISM  Mouse mammary tumor virus
                 Viruses; Retroid viruses; Retroviridae; Betaretrovirus; Mammalian
                 type B retroviruses.
     REFERENCE   1  (bases 1 to 10125)
       AUTHORS   Moore,R., Dixon,M., Smith,R.E., Peters,G. and Dickson,C.
       TITLE     Complete nucleotide sequence of a milk transmitted mouse mammary
                 tumor virus: Two frameshift suppression events are required for
                 translation of gag and pol
       JOURNAL   J. Virol. 61, 480-490 (1987)
      MEDLINE    87112944
     COMMENT     Draft entry and computer readable copy of sequence in [1] were
                 kindly provided by C.Dickson, 05-MAY 1987.
                 The exact frameshift sites are undetermined and were annotated as
                 introns at the 5' most possible site.
     FEATURES             Location/Qualifiers
          source          1..10125
                          /organism="Mouse mammary tumor virus"
                          /db_xref="taxon:11757"
          repeat_region   120..125
                          /note "target site duplication"
          LTR             126..1453
                          /note "5' LTR"
          prim_transcript 1321..8750
                          /note "env mRNA"
```

```
intron          1699. .6682
                /note "env intron A"
CDS             1633. .3408
                /note="Pr72 gag precursor"
                /codon_start =1
                /protein_id="AAA46543.1"
                /db_xref="GI:332130"
                /translation="MGVSGSERQPFEVSVEQELRSRLELFKQDPSVGCKDIIGVLYLPSSQTLQHLYAETLEVYQEIEKVSP
                WFPEEGGLNLQDWKRVGREMKRYAAEIKRCDIPDSIPKQAYPIWLQSLPEILIEQGSDLVLLSA
                EAKSVTEPFLPEGLDGLLSTSSQFFYGYRIGIAYAEFIDIEVDKLSEHIYDEPYREKEEK
                ADKNEEKDHVRKIKKVVQRKENSEGKKKEKDSKAFLATDWNDDDLSPEDWDDLEEQAA
                HYHDDDEILLPVKEKVVKKEDPQALRRFLDTVGFACAMALAREKGDLDLTFTPVVVPCE
                SDEDDTPVWEPLPLKILKELQSAVRTMGDSAPYTLQVVDMVASQWLIPSDWHQTARAT
                LSPGDYVLWRTLYFEKSKFMVQKAAGNKKGESVSLDMLIGCDQSLESPSSQTRLSRLQVLR
                DVFTNAVLAWRAIPPPGVKKTVLAGLKQGNEESYFTFISRLEEAVYRMMDPRGEGSDILL
                IKQLAWENANSLQQDLIRFLRKTGTLQDVIRACLDASPAVVQGMAYAAAMRQGKYSTF
                VKQTYGGGKGQGAPGVCFSCGKTGHLERRCPEKDRQKCHKCGKIGHWAKSE
                CKSKFDKDGNPLPPLETNAENSKNL"
exon            <1633. .3387
                /note "Pr120 gag pro precursor"
                /number 1
CDS             join(1633. .3387,3390. .4196,4199. .6898)
                /note="Pr160 gag pro pol precursor"
                /codon_start =1
                /protein_id "AAA46542.1"
                /db_xref "GI:332129"
                /translation="MGVSGSERQPFEVSVEQELRSRLELFKQDPSVGCKDIIGVLYLPSSQTLQHLYAETLEVYQEIEKVSP
                WFPEEGGLNLQDWFRVTEPMERYAAEIKRCDIPDSIPKQAYPIWLQSLPEILIEQGSDLVLLSA
                EAKSVTEPFLPEGLDGLLSTSSQFFYGYRIGIAYAEFIDIEVDKLSEHIYDEPYREKEEK
                ADKNEELKIHVEKIKKVVQRKENSEGKKKEKDSKAFLATLGNDDDLSPEDWDDLEEQAA
                HYHDDDFILLPVKEKVVKKEDPQALRRFLDTVGFACAMALAREKGDLDLTFTPVVVPCE
                SDEDDTPVWEPLPLKILKELQSAVRTMGDSAPYTLQVVDMVASQWLIPSDWHQTARAT
                LSPGDYVLWRTLYFEKSKFMVQKAAGNKKGESVSLDMLIGCDQSLESPSSQTRLSRLQVLR
                DVFTNAVLAWRAIPPPGVKKTVLAGLKQGNEESYFTFISRLEEAVYRMMDPRGEGSDILL
                IKQLAWENANSLQQDLIRFLRKTGTLQDVIRACLDASPAVVQGMAYAAAMRQGKYSTF
                VKQTYGGGKGQGAPGVCFSCGKTGHLERRCPEKDRQKCHKCGKIGHWAKSE
                CKSKFDKDGNPLPPLETNAENSKNLEQVSLSFALKSGVDVSKYVTQNIGVCVTVLSLQTLT
                GTPGSAGLDLESSQKDLIEEIQKVSLVLEDSVSLVELEGEEIRLKSAYNPKDPLFAWLYG
                GVIDSDFQGGETKVMVKAAENAVSIMSCERRIAQLELLLPYLRLPNPVIKEEEKGSEGEQST
                SHVHWVQETSDSKFMLHIYLNGRRTLGLLETGADRTIAIRNAEFANWPTHQTESSIQG
                LGMACGVARSSQPLEPWQHEDKSCTIRFFVTLFLPEIIWGRDIMKHIEFVKLMTDSPLDS
                GFMIGAIESNLEPADQIGWKSQQPVWLNQWPLPQGKKIQALVQLVTEQLQLGHIEEGNSF
                WNTPVFVIKKKSGKWRLLQDLRAVNAIMQMGALQPGLPSPVAVEEGWELIIIDLQDC
                FFNIKLHPTDQKFAFTLPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFRIQN
                DFYQDSYIVHYMDDILIAHPDKSCVEFLTQMIAELNKEGYEVSWKKLQEKIVNLGQEVEH
                THQGDSVSYQKIQIPNTDKLPITNFQLELIGNINWIRPNLAPDLSHLTAPIRCLLKKDSK
                NDISTRKLTPEAQKAIQLMNEKLQTAPVLRLPDFSKPFYTFVDASLWGVGCVV
                EWIHLPHISEKVIEEVNKGTKQIIGYASRQLKEHEKNYPVHDLELAAVIFALKIWRHYLYG
                EDWFISTDCRSLQWLFNFRNPNGKLARWALLLQEFDFKIEYRPGKDNVVADALSRIEVGD
                SVTYIQGKEPIIKENTQNTAQQAEIVAVITAFEEVSQPENLYTDSKYVTGIFFIGKIA
                TLSPRTKIYIELKHLQELIHKRQFEYFQHTRQHEQLPCPIAQQNAYADSLTRTLTAL
                ESAQESHATHHQNAAAIRFQFPITRPQAREFIVKLCPNCPIWGHAPQLGVNPPGCLKPDV
                LWQMDVTHVSEFGKIKYVHVIVDTYSHFIFAIAKDGEAIEVVQHLAQSFAYMGIPQK
                IKTDNAPAYVSKRSIQEFLAPWFKCHVDGPYNPQGQAIVERTHQNIEAQTNELQEAGE
                YYTPHHLLAHALFVLNHVNMDNQGHTAAERHWCPISADPKIMVMWREQLLIGCWFQPVV
                LITAGRDCYACVEPQDAPTPIWVPDPRFLFPEFPRKFAIPIPDIAEKIPPPDEKDQQESP
                KNESSPHQREDGLATSAGVDLRSGGP"
CDS             join(1633. .3387,3390. .4214)
                /note="Pr110 gag pro precursor"
                /codon_start =1
                /protein_id="AAA46541.1"
                /db_xref="GI:332128"
                /translation="MGVSGSERQPFEVSVEQELRSRLELFKQDPSVGCKDIIGVLYLPSSQTLQHLYAETLEVYQEIEKVSP
                WFPEEGGLNLQDWFRVGPEMKRYAAEIKRCDIPDSIPKQAYPIWLQSLPEILIEQGSDLVLLSA
                EAKSVTEPFLPEGLDGLLSTSSQFFYGYRIGIAYAEFIDIEVDKLSEHIYDEPYREKEEK
                ADKNEEKDHVRKIKKVVQRKENSEGKKKEKDSKAFLATDWNDDDLSPEDWDDLEEQAA
                HYHLDDELILLPVKEKVVKKEDLQALRRFLDTVGFACAMALAREKGDLDLTFTPVVVPCE
                SDEDDTPVWEPLPLKILKELQSAVRTMGDSAPYTLQVVDMVASQWLIPSDWHQTARAT
                LSPGDYVLWRTLYFEKSKFMVQKAAGNKKGESVSLDMLIGCDQSLESPSSQTRLSRLQVLR
                DVFTNAVLAWRAIPPPGVKKTVLAGLKQGNEESYFTFISRLEEAVYRMMDPRGEGSDILL
                IKQLAWENANSLQQDLIRFLRKTGTLQDVIRACLDASPAVVQGMAYAAAMRQGKYSTF
                VKQTYGGGKGQGAPGVCFSCGKTGHLERRCPEKDRQKCHKCGKIGHWAKSE
                CKSEEDKIGNPLPPLETNAEPKEELKSLTETLEDQLSPEFQTWLPETENSENFRDIQKLE
                QKEQLFPQWQELSMASKEHKARLLAGAEDSIIQLVQEAQTAEKSPFIQDRLAGLNERVQ
                QMSLLATVQIAITALEQQHTVQPPNPEVEKTAKDQFEKKLNTLKTVRTKATEKERQRLVE
                DKEKEFYFLGKLEDLVSQSEHFHTSSGVKERAQKLYSRLGAIWEDVGKQLTDYGQPMVR
                VLREQIEKEFETMEKLISDIDELERKAAQLEAMCTLAQPLPGPLEKRAGGLLESLKSVSL
                LHVMRKLSTHVYKDYPEEVLDPSVKDLIDQLLMLKKDQRLNREANDLMNQNRPWVGDIP
                KGPKEFEDYVDRFFKALRAEQATQEVKNWMTETLLVQNANPDCKTILRALGPGATLEEMM
                TACQGVGGPGHKARVLAEAMSQVTNSATIMMQRGNFRNQRKIVKCFNCGKEGHIAKNCR
                APRKKGCWKCGKEGHQMKDCTERQANFLGKIWPSYKGRPGNFLQSRPEPTAPPEESFRF
                GEETTTPSQKQEPIDKELYPLASLRSLFGSDPSSQ"
exon            1633. .3387
                /note "Pr120 gag pro precursor"
                /number 1
exon            3390. .4196
                /note "Pr160 gag pro pol precursor"
                /number 2
exon            4199. .6898
                /note "Pr160 gag pro pol precursor"
                /number 3
CDS             6684. .8750
                /note "env precursor"
                /codon_start =1
                /protein_id "AAA46544.1"
                /db_xref "GI:332131"
                /translation="MRVMGIQRNCQHLWRWGTMILGMLMICSAAEQLWVTVYYGVPVW
                KEATTTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVVLENVTENFNMWKNDMVDQMHE
                DIISLWDQSLKPCVKLTPLCVTLNCTDLRNVTNATNSSSGRMIMEKGEIKNCSFNISTSI
                RGKVQKEYAFFYKLDIIPIDNDTTSYKLTSCNTSVITQACPKVSFEPIPIHYCAPAGFAI
                LKCNDKKFNGTGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEEVVIRSVNFTDNAKTII
                VQLNESVEINCTRPNNNTRKSIRIGPGQAFYATGDIIGDIRQAHCNISRAKWNNTLKQIV
                SKLREQFGNNKTIIFKQSSGGDPEIVTHSFNCGGEFFYCNSTQLFNSTWFNSTWSTEGSN
                NTEGSDTITLPCRIKQIINMWQKVGKAMYAPPISGQIRCSSNITGLLLTRDGGNSNNESE
                IFRPGGGDMRDNWRSELYKYKVVKIEPLGVAPTKAKRRVVQREKRAVGIGALFLGFLGAA
                GSTMGAASITLTVQARQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAVERYL
                KDQQLLGIWGCSGKLICTTAVPWNASWSNKSLEQIWNHTTWMEWDREINNYTSLIHSLIE
                ESQNQQEKNEQELLELDKWASLWNWFNITNWLWYIKLFIMIVGGLVGLRIVFAVLSIVNR
                VRQGYSPLSFQTHLPTPRGPDRPEGIEEEGGERDRDRSIRLVNGFLALAWDDLRSLCLFS
                YHRLRDFILIAARTVELLGRSSLKGLQRGWEILKYLGSLVQYWGLELKKSAISLLDTIAI
                AVAEGTDRIIEFIQRICRAILHIPTRIRQGFERALL
exon            6898. .7660
                /note "Pr73 precursor"
                /codon_start =1
                /protein_id "AAA46545.1"
                /db_xref "GI:332132"
                /translation="MPEPLQPWLNGSLTFLSGAAPLLTSSFNGSSVQLIYNLNFSIDSIFLLHQGP
                LFISSYKFHQPLPNPRGKTNFLGYIFSLCPDWIMTKHGTLFVGLPEYTYKFHLGIYSKK
                IIAFLGFYSCGYPLGLNDGWVVNGYKVLYKSLSPLSLAVALGLFWFMLPLTLSPNYQRW
                VDQYTVSLFGMSSIWGKIFDTLEKRTVAALLIGHYSAKIYGMSYYD"
exon            8699. .10026
                /note "3' LTR"
repeat_region   10027. .10032
                /note "target site duplication"
BASE COUNT      2929 a  2149 c  2298 g  2668 t
ORIGIN      138 bp upstream of PstI site.


alignment scores:
                Quality:    9.00              Length:      9
                  Ratio:    1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_75_83 & MMTPROCG  ..

 Align seq 1/1  to: MMTPROCG  from: 1  to: 10125

        1 AlaLysArgArgValGlyAlaSerSerArg 9
          ||||||||||||||||||||||||||||||||||
    10069 GCCAAAGCCCCCGGGTGCCTCTCGCCGG 10095

Seq_name: gb_UL:HSVACG074221
```

```
Seq documentation block:
LOCUS       AC074221    117655 bp    DNA              HTG       15 NOV 2000
DEFINITION  Mus musculus chromosome 15 clone RP23 81A4 strain C57BL6/J, WORKING
            DRAFT SEQUENCE, 35 unordered pieces.
ACCESSION   AC074221
VERSION     AC074221.5  GI:11177936
KEYWORDS    HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorphi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 117655)
  AUTHORS   Han,J., Montgomery,K.T., Grills,G., Lee,E., Long,J., Pomerantz,R.,
            Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
            Gordon,M., Goltz,J.S. and Kucherlapati,R.
  TITLE     High Throughput Mouse Sequencing
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 117655)
  AUTHORS   Han,J., Montgomery,K.T., Grills,G., Lee,E., Long,J., Pomerantz,R.,
            Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
            Gordon,M., Goltz,J.S. and Kucherlapati,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20 JUL 2000) Department of Molecular Genetics, Albert
            Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
            Bronx, NY 10461, USA
COMMENT     On Nov 15, 2000 this sequence version replaced gi:11067106.
                  --  Genome Center
            Center:  Albert Einstein College of Medicine
            Center Code:  AECOM
            Web site:
            http://sequence.aecom.yu.edu/gsi
            bin/ws-cgi/mouse00k/mouseSEQ/mouseseqtable.hts
            Contact:  jhan@sequence.aecom.yu.edu
                    -Summary Statistics
             Center project name:  ADZ
            Sequencing vector:  pUC18;  L08752
            Chemistry:  Dyeterminator Big Dye;  100%
            *Consensus quality:  109029 at least Q20
            *Consensus quality:  105812 at least Q30
            *Consensus quality:  99196 at least Q40
            Estimated insert size:  approse FP   N/A
            **Estimated insert size: 116975  sum of contigs
            Quality coverage:  approse FP   N/A
            Quality coverage:  6.2x sum of contigs    N/A
                    --
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 35 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     11545: contig of 11545 bp in length
            *   11546     11555: gap of unknown length
            *   11556     20397: contig of 8842 bp in length
            *   20398     20417: gap of unknown length
            *   20418     28110: contig of 7693 bp in length
            *   28111     28130: gap of unknown length
            *   28131     35775: contig of 7645 bp in length
            *   35776     35795: gap of unknown length
            *   35796     39098: contig of 3303 bp in length
            *   39099     39118: gap of unknown length
            *   39119     45564: contig of 6446 bp in length
            *   45565     45584: gap of unknown length
            *   45585     50931: contig of 5347 bp in length
            *   50932     50951: gap of unknown length
            *   50952     55271: contig of 4320 bp in length
            *   55272     55291: gap of unknown length
            *   55292     59087: contig of 3796 bp in length
            *   59088     59107: gap of unknown length
            *   59108     63969: contig of 4862 bp in length
            *   63970     63989: gap of unknown length
            *   63990     67991: contig of 4002 bp in length
            *   67992     68011: gap of unknown length
            *   68012     71273: contig of 3262 bp in length
            *   71274     71293: gap of unknown length
            *   71294     74451: contig of 3058 bp in length
            *   74452     74471: gap of unknown length
            *   74472     78263: contig of 3892 bp in length
            *   78264     78283: gap of unknown length
            *   78284     81256: contig of 2973 bp in length
            *   81257     81276: gap of unknown length
            *   81277     83650: contig of 2374 bp in length
            *   83651     83670: gap of unknown length
            *   83671     86224: contig of 2554 bp in length
            *   86225     86244: gap of unknown length
            *   86245     88380: contig of 2136 bp in length
            *   88381     88400: gap of unknown length
            *   88401     91500: contig of 3100 bp in length
            *   91501     91520: gap of unknown length
            *   91521     93802: contig of 2282 bp in length
            *   93803     93822: gap of unknown length
            *   93823     95981: contig of 2159 bp in length
            *   95982     96001: gap of unknown length
            *   96002     98819: contig of 2818 bp in length
            *   98820     98839: gap of unknown length
            *   98840    100597: contig of 1758 bp in length
            *  100598    100617: gap of unknown length
            *  100618    102191: contig of 1574 bp in length
            *  102192    102211: gap of unknown length
            *  102212    104013: contig of 1802 bp in length
            *  104014    104033: gap of unknown length
            *  104034    105993: contig of 1960 bp in length
            *  105994    106013: gap of unknown length
            *  106014    106894: contig of 881 bp in length
            *  106895    106914: gap of unknown length
            *  106915    108469: contig of 1555 bp in length
            *  108470    108489: gap of unknown length
            *  108490    109399: contig of 910 bp in length
            *  109400    109419: gap of unknown length
            *  109420    111041: contig of 1622 bp in length
            *  111042    111061: gap of unknown length
            *  111062    113148: contig of 2087 bp in length
            *  113149    113168: gap of unknown length
            *  113169    114596: contig of 1428 bp in length
            *  114597    114616: gap of unknown length
            *  114617    115542: contig of 926 bp in length
            *  115543    115562: gap of unknown length
            *  115563    117012: contig of 1450 bp in length
            *  117013    117032: gap of unknown length
            *  117033    117655: contig of 623 bp in length.
FEATURES             Location/Qualifiers
     source          1..117655
                     /organism "Mus musculus"
                     /strain "C57BL6/J"
                     /db_xref "taxon:10090"
                     /chromosome "15"
                     /clone "RP23 81A4"
                     /sex "male"
BASE COUNT     27336 a  32198 c  41011 g  26462 t     238 others
ORIGIN


alignment scores:
                Quality:     9.00              Length:       9
                  Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 92 99 120 94 P 2 2 PY 2 94 x AC074221/rev   ..

 Align seq 1/1  to reverse of: AC074221  from: 1  to: 117655

       1 AlaLysAlaProGlyAlaSerSerArg 9
```

```
          ||| ||||||||||||||||||||||||
    45021 GCCAAAGCCCCCGGTGCTTCGTCCCGG 34995

seq_name: qb_pr6:AL157472

seq_documentation_block:
LOCUS       AL157472    150660 bp    DNA              PRI       29-NOV-2000
DEFINITION  Human DNA sequence from clone RP11-375E1 on chromosome 6, complete
            sequence.
ACCESSION   AL157472
VERSION     AL157472.19  GI:11494244
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo
REFERENCE   1  (bases 1 to 150660)
  AUTHORS   Skuce,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29 NOV 2000) Sanger Centre, Hinxton, Cambridge,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 30, 2000 this sequence version replaced gi:11342772.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows.  An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements.  Where the sequence is
            ambiguous, there is an annotation using the 'unsure' feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:  EMBL; Sw:  SWISSPROT; Tr:  TREMBL; Wp:  WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-375E1 is from the library RPCI 11.2 constructed at the Roswell
            Park Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-375E1. It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP11-375E1 is at 1 in this sequence. The
            true left end of clone RP3-391022 is at 150561 in this sequence.
            The true right end of clone RP11-302L10 is at 113044 in this
            sequence.
FEATURES             Location/Qualifiers
     source          1..150660
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-375E1"
                     /clone_lib="RPCI-11.2"
BASE COUNT     45470 a   98887 c   88740 g   47563 t
ORIGIN


alignment_scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_75_83 x AL157472.19:
```

```
 Align Seq 1/1  to reverse of: AL157472  from: 1  to: 150660

       1 AlaLysAlaProGlyAlaSerSerArg 9
         ||||:|||||:|||:|||||||||:|||
  129518 GCCAAGGCCCCTGGGACCAGCAGTGGG 129492

seq_name: qb_bal:AE005854

seq_documentation_block:
LOCUS       AE005854    11094 bp    DNA              BCT       28-MAR-2001
DEFINITION  Caulobacter crescentus  section 180 of  359 of the complete genome.
ACCESSION   AE005854 AE005673
VERSION     AE005854.1  GI:13423223
KEYWORDS    .
SOURCE      Caulobacter crescentus.
  ORGANISM  Caulobacter crescentus
            Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
            Caulobacter.
REFERENCE   1  (bases 1 to 11094)
  AUTHORS   Nierman,W.C., Feldblyum,T.V., Paulsen,I.T., Nelson,K.E., Eisen,J.,
            Heidelberg,J.F., Alley,M.R.K., Ohta,N., Maddock,J.R., Potocka,I.,
            Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D., Ely,B., DeBoy,R.T.,
            Dodson,R.J., Durkin,A.S., Gwinn,M.L., Haft,D.H., Kolonay,J.F.,
            Smit,J., Craven,M.B., Khouri,H., Shetty,J., Berry,K., Utterback,T.,
            Tran,K., Wolf,A., Vamathevan,J., Ermolaeva,M., White,O., Salzberg,S.L.,
            Venter,J.C., Shapiro,L. and Fraser,C.M.
  TITLE     Complete genome sequence of Caulobacter crescentus
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)
  MEDLINE   21178698
REFERENCE   2  (bases 1 to 11094)
  AUTHORS   Nierman,W.C., Feldblyum,T.V., Paulsen,I.T., Nelson,K.E., Eisen,J.,
            Heidelberg,J.F., Alley,M.R.K., Ohta,N., Maddock,J.R., Potocka,I.,
            Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D., Ely,B.,
            DeBoy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L., Haft,D.H.,
            Kolonay,J.F., Smit,J., Craven,M., Khouri,H., Shetty,J., Berry,K.,
            Utterback,T., Tran,K., Wolf,A., Vamathevan,J., Ermolaeva,M.,
            White,O., Salzberg,S.L., Shapiro,L., Venter,J.C. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21 JAN 2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1..11094
                     /organism="Caulobacter crescentus"
                     /db_xref="taxon:caucr"
     gene            complement(103..1257)
                     /gene="CC1792"
     CDS             complement(103..1257)
                     /gene="CC1792"
                     /note="identified by match to protein family HMM"
                     /codon_start=1
                     /transl_table=11
                     /product="ROK family protein"
                     /protein_id="AAK23268.1"
                     /db_xref="GI:13424224"
                     /translation="MPASKSERSQDRVSKGEKLRAGIDVGRVSVKAFLA
                     ARKIGLAPLSVIKLIADLLDRDLVIEGARVVAGRGQVSLFLSLARDAAESTGVSLIFD
                     ALSLALIDLSGVVRDAWREPFETEDKSAALIRLEKIFDLLATRAAIDKSRLLGVGVAA
                     SGEFTGLKRQLNLPLSMEQWALSDLDAFESARIGLDTIWTDNDGNAAAAGEALYGVGKT
                     LSSFAYLYVGRGLGGGVVLDHSLWRGRNGNAGEFTGLLSPDRRADKPIDGLLKEMFAA
                     DKAFASIDQMLAELDPASPVIEAWAKRIAEELNALLSAIAAIIDTEAVVIGGKAIPA
                     LVDALIGQASYYSVPIKDRDRIFPIILRSQVVRDVIAIGAASMPFSGHEF"
     gene            complement(1269..1475)
                     /gene="CC1793"
     CDS             complement(1269..1475)
                     /gene="CC1793"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAK23269.1"
                     /db_xref="GI:13424225"
```

/translation="MNRHALgx[illegible]PAQPLGLGAQKPYRKAASAAMMV"
gene 1464..1559
/gene="CC1794"
CDS 1464..1559
/gene="CC1794"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK23773.1"
/db_xref="GI:13423226"
/translation="MAAHCAPLSGPLLSGSGLAMLGPLGVPALN"
gene 1564..1797
/gene="CC1795"
CDS 1564..1797
/gene="CC1795"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK23774.1"
/db_xref="GI:13423227"
/translation="MKRATRSGAAEAYVLGKRMKEVSLDPESLLERRDKEFWTGLLED
DEAELRVLQTQFDPHPLAVEDALKAHRCHAQAR"
gene complement(2091..3536)
/gene="CC1796"
CDS complement(2091..3536)
/gene="CC1796"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK23775.1"
/db_xref="GI:13423228"
/translation="MDLSRFVEGCDRRQPTLLPECLDDYVAEENPVRVVDVFVDDLDL
GTLGFEGLTPAVTGRPAYHPAMLLKLYVYGYLNKVQSSRKLERECARRNVELMWLLGKL
APDHKTIADFRKDNGPAIQAACAQFVVLCPQLGLFGAAIVAIDGSKFKAVNNRDRNFT
ANKLKKKLEQVAERIAGYLRDLDTADRQEGEAAFTKSFKLKEKVERLREQMRWLQTME
AKVAAGEDQQVSLTEPDARSMATSHKGVGYNVQVAVETKHHLIVAHEVVNQGSDR
QQLSSMAARTKDAIGVETLLVDLADRGYFSGKELLACEALGVIPYVPRPLTSGAKANGR
FGKQDFVYLPEQDVYRCPSGALLPRHMTNVENGMTLHRYWDHPSCQSCALKFQCTTSS
ARRVTRWEHEAVIDAMQRRMDLFPGAMRQRPETVEHPFGFLKAWMGSTHFLMKRLFNV
KTEIGLHILAYNLKRVIAILGVGELMKATQA"
gene complement(4276..4279)
/gene="CC1797"
CDS complement(4276..4279)
/gene="CC1797"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK23776.1"
/db_xref="GI:13423229"
/translation="MMGMPSPPPVPSPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
RGRYPWWQLGAGIGARKSCVDSPAAGAIGARAWEEEEEEEEEEEGNGVLSFFACLVA
LVEGTIETRRAESHQPPTVPPPMPGNADACPESQAPSGPSQLPHMGPNAPSSPIAHVPI
VLLSTTY"
gene 4161..5406
/gene="CC1798"
CDS 4161..5406
/gene="CC1798"
/note="identified by match to PFAM protein family HMM
PF00441"
/codon_start=1
/transl_table=11
/product="acyl-CoA dehydrogenase"
/protein_id="AAK23777.1"
/db_xref="GI:13423230"
/translation="MTDPDTLSALIDVIQRFVAERLRPIEGLVSEIDEVPGSLIEEMK
QLGLFGLSIPEEYGGLGLSLLEAVIVAEEIAKVIVALSVAFGAHSGIGVNGLVLVHESDE
AQKARWLPSIASGETIAFALTEAEAGSDSASVQTRAVRDGDHYVLNGVKRYITNAGR
ANLFTVMARTDPNTKDGAVSAFLVPADLPGLSVGKPEKKMGQQGAPTHEVVFEDVRV

[illegible]
EQETQAMTAQSKTEALAAKALVEDGARSDGSANVTEAAAIKLEASEMVGRVADEAV
QVFGGAGYVADYGIERLYRDVRIFRIYEGASQIQQLIIAREILKRGG"
gene complement(5369..7039)
/gene="CC1799"
CDS complement(5369..7039)
/gene="CC1799"
/note="identified by match to PFAM protein family HMM
PF00743"
/codon_start=1
/transl_table=11
/product="steroid monooxygenase"
/protein_id="AAK23778.1"
/db_xref="GI:13423231"
/translation="MTASQQPTASQDPTETDALVIGAGFGGMYAVHKLRDMGLSVQGF
EAGGDVGGVWYWNRYPGARCDIESIDYCYSFSPDLFQEWTWSEQFAAQSELLAYANFV
ADRLDLRRHFLFETWVTGAAYDEAANVWRVTVDDGRVFEARYCVMASGPLSTPKGVPE
IPGAEEFKGEILHAAKWPHTPVSFEGKRVGLVGTGSTGIQVIPEVAKTAGALTVFQRTP
SFTLPMRNVTLEPDYVAELKKNYAGIRQVARNSFAGVRFQESTRPFESVTPEQRRALM
EDSWAAGGLSFLGTFSDLLVNPEANDFVAEFVRSKIRSIVKDPVTAEKLKPKGYPIFA
RRPCLDTDYYETFNLPHVKLHDCLESPIVKLTPAGLLTEDGVEYDVLILATGYDGLT
GALSNIDVVGRGGLQLPEKWRDGAESYLGMTASGFPNLFTVTGPGSPSVLANLIVSIE
QHIEWIADAITHMRQEGLATIEPTPEAEQEWMDMVYGMAETLVPKANTWYVGGNISG
KPRVFLSMYIGGFQKYSEACPLAAEKGYEGFVFEKAATKQGARSPLLVG"
gene 7127..8098
/gene="CC1800"
CDS 7127..8098
/gene="CC1800"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK23779.1"
/db_xref="GI:13423232"
/translation="MCLLAVEAGAPLTESPGRRMAEPGAIWPWFGRELEEQMTDQNAS
TTPQGASPSRLMKFGATPGSSRQTLIADRRLCVGPPGNVFVFGGVLFGSALAALEDHL
QKPVVWSTIQFLGFVRAEDVLALDIEVPSPGRTISQARVIGRVGDRETFLASAAIGVR
DEPLFHAWDQAPEVPPPEACETAKLWPHGQEPGTLLETLEMPTVPGAFAEVEEPTGSLA
DSGKTVFWVRAKQGLPMDTSVLALFGDETPYATSKALGGGLGGVSLDNTLKVLRKVET
IAGIGTAKIQGIGRGIFAIGEIKLLAKTGELMAVGSQGMNAGFEFAKA"
gene complement(8102..9787)
/gene="CC1801"
CDS complement(8102..9787)
/gene="CC1801"
/note="identified by match to protein family HMM"
/codon_start=1
/transl_table=11
/product="4-coumarate--CoA ligase, putative"
/protein_id="AAK23777.1"
/db_xref="GI:13423233"
/translation="MIYRPLLGPGAAFIPFPHRFLAERFPAAVPWNTGPTTDAVARA
AFEEEALIWRPVVVIWSETDIKTNNMAKALIPGLKFTAKTTLGFNHFAYLEFLVA
GVKAKGLTVNGNYKYTADLIAYVLNGNLNALVLYQGSFLGMLEPLKAPLPGVQAWVRT
PNLPGPLAAPFGALSYFTLAPTGPGPFLPSAPGWSYLTYPSLGPPVPVHWSTPQ
CPPVVLPLTPLAPVPLPMLPHLPTVPAVPPGPAPLPMPHPPPPGNGAMGLLNGGIAT
LLSGRPPAVQLWEPVPFPGATFVSTVGLVLAFLMLFALGLAPGAGPLFSMPATSSAG
LWEKLVKLGLFHMPNTALVSTLGALLAGPLYLVAPAGPFPLLSGVFLAAPLIVLLL
PPPGAVMPPPGSGPLLAPGSVMANGVPPDPLKLAALYKLLDNVRYAVPGGLARWMPP
NQPTLIGRGNLSINTGGPKVFPKEVKEALKLQDPGVADVLVVGQPPDPKKWSKIVVATLKA
PAVFPEAQVPKAGLAQTLSAYKHPKRFVLVDIPKPHESGKADYRIATALAAELA"
gene 9734..11026
/gene="CC1802"
CDS 9734..11026

alignment scores:
quality: 8.00 length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment blocks:
85 89 126 94 95 2 5 6 Y 75 83 x AE005854 ..

Align Seq 1/1 to: AE005854 from: 1 to: 11094

```
      2 LysAlaProGlyAlaSerSerArg 9
        |||||||||||||||||||||||||
   8823 AAGGCGCCCGGCGCGTCGTCCACG 8846

seq_name: qb_htg10:AC025796

seq_documentation_block:
LOCUS       AC025796   176612 bp    DNA              HTG       04 APR 2000
DEFINITION  Homo sapiens clone RP11-63818, WORKING DRAFT SEQUENCE, 18 unordered
            pieces.
ACCESSION   AC025796
VERSION     AC025796.2  GI:7408048
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176612)
  AUTHORS   Kirren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-63818
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176612)
  AUTHORS   Kirren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,E.,
            Bogoslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choopel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Hoaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,P., Landers,T., Lehoczky,J.,
            Levine,E., Lieu,C., Liu,C., Lorke,K., Macdonald,P., Margolis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPherters,R.,
            Meldrim,J., Meneus,L., Mihova,I., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,E., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schaper,S., Severy,R., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Testave,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,R., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 420 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 4, 2000 this sequence version replaced gi:7239682.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ---------- Genome Center
                Center: Whitehead Institute/MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            ---- --------- Project Information
                Center project name: L8157
                Center clone name: 648_I_8
            ---------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 167495 bases at least Q40
                Consensus quality: 172009 bases at least Q30
                Consensus quality: 173835 bases at least Q20
                Insert size: 184000; agarose fp
                Insert size: 174912; sum-of-contigs
                Quality coverage: 4.4 in Q40 bases; agarose fp
                Quality coverage: 4.6 in Q20 bases; sum-of-contigs
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 18 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1071: contig of 1071 bp in length
            *     1072  1171: gap of      100 bp
            *     1172     2854: contig of 1683 bp in length
            *     2855  2954: gap of      100 bp
            *     2955     4616: contig of 1662 bp in length
            *     4617  4716: gap of      100 bp
            *     4717     6907: contig of 2191 bp in length
            *     6908  7007: gap of      100 bp
            *     7008     8967: contig of 1960 bp in length
            *     8968  9067: gap of      100 bp
            *     9068    11248: contig of 2181 bp in length
            *    11249 11348: gap of      100 bp
            *    11349    14274: contig of 2926 bp in length
            *    14275 14374: gap of      100 bp
            *    14375    16941: contig of 2567 bp in length
            *    16942 17041: gap of      100 bp
            *    17042    20895: contig of 3854 bp in length
            *    20896 20995: gap of      100 bp
            *    20996    25731: contig of 4736 bp in length
            *    25732 25831: gap of      100 bp
            *    25832    32760: contig of 6929 bp in length
            *    32761 32860: gap of      100 bp
            *    32861    43355: contig of 10495 bp in length
            *    43356 43455: gap of      100 bp
            *    43456    50896: contig of 7441 bp in length
            *    50897 50996: gap of      100 bp
            *    50997    62663: contig of 11667 bp in length
            *    62664 62763: gap of      100 bp
            *    62764    76576: contig of 13813 bp in length
            *    76577 76676: gap of      100 bp
            *    76677    94726: contig of 18050 bp in length
            *    94727 94826: gap of      100 bp
            *    94827   112911: contig of 18085 bp in length
            *   112912 113011: gap of      100 bp
            *   113012   176612: contig of 63601 bp in length.
FEATURES             Location/Qualifiers
     source          1. .176612
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-63818"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1071
                     /note "assembly_fragment"
     misc_feature    1172. .2854
                     /note "assembly_fragment"
     misc_feature    2955. .4616
                     /note "assembly_fragment"
     misc_feature    4717. .6907
                     /note "assembly_fragment"
     misc_feature    7008. .8967
                     /note "assembly_fragment"
     misc_feature    9068. .11248
                     /note "assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    11349. .14274
                     /note "assembly_fragment"
     misc_feature    14375. .16941
                     /note "assembly_fragment"
     misc_feature    17042. .20895
                     /note "assembly_fragment"
     misc_feature    20996. .25731
                     /note "assembly_fragment"
     misc_feature    25832. .32760
                     /note "assembly_fragment"
     misc_feature    32861. .43355
                     /note "assembly_fragment"
```

```
     misc_feature    43456..50896
                     /note "assembly fragment
                     clone end:SP6
                     vector side:right"
     misc_feature    50997..62663
                     /note "assembly fragment"
     misc_feature    62764..76576
                     /note "assembly fragment"
     misc_feature    76677..94726
                     /note "assembly fragment"
     misc_feature    94827..112911
                     /note "assembly fragment"
     misc_feature    113012..176612
                     /note "assembly fragment"
BASE COUNT    49840 a   48680 c   49896 g   46505 t    1701 others
ORIGIN


Alignment scores:
              Quality:     8.00              Length:       8
                Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

Alignment block:
 US-09-126-945B-2-COPY_75_83 x AC025796   ..

 Align seq 1/1   to: AC025796   from: 1   to: 176612

        1 AlaLysAlaProGlyAlaSerSer 8
          ||||||||||||||||||||||||
   107090 GCGAAGCCCCCGGCCGCAICTICA 107113
```

```
OM of: US-09-126-945B-2_COPY_84_92 to: GenEmbl:*   ent_format . pts

Date: Sep 14, 2001  9:47 PM

About: Results were produced by the GenCore software, version 4.5.
       Copyright (c) 1993 2000 Compaqen Ltd.

Command line parameters:
 -MODEL frames_p2n.model -DEV xip
 -Q "res_1 SPI_sp*.fas:gi_6:5 format 11 Q:\cgn_jobs\11_17_01\pii_poly_tail_1_242
 -DB GenEmbl -QFMT fastap -SUFFIX olip2n.rge -GAPOP 4.500
 -GAPEXT 0.050 -MINMATCH=0.100  LOOPCL 0.000 -LOOPEXT 0.000
 -QGAPOP 4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT 60.000
 -FGAPOP 6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT 60.000
 -DELOP 6.000 -DELEXT 7.000 -START 1 -MATRIX oligo
 -TRANS human40.cdi -LIST 45 -DOCALIGN 200 -THR_SCORE quality
 -THR_MIN=1 -ALIGN 15 -MODE LOCAL -OUTFMT pfs  NORM ext  MINLEN 0
 -MAXLEN 2000000000 -USER US09126945B_COPY_1 7594  NCPU 6
 -NCPU 6 -LONGLOG -NO_XLPXY -WAIT -THREADS 1

Search Information block:
Query: US-09-126-945B-2_COPY_84_92
Query length: 9
Database: GenEmbl:*
Database sequences: 1444157
Database length: -856060004
Search time (sec): 6503.700000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000


score_list:
Sequence              Strd orig    ZScore    EScore Len  ! Documentation ..
qb_pat1:AX048148       +     9.00   173.13   0.2088    1894  ! AX048148 Sequence 1 from Patent
qb_pr4:AF071538        +     9.00   173.13   0.2088    1894  ! AF071538 Homo sapiens Ets tran
qb_pr1:AB031549        +     9.00   173.12   0.2091    1898  ! AB031549 Homo sapiens PSF mRNA
qb_pat1:AX048154       +     9.00   169.67   0.3256    3417  ! AX048154 Sequence 12 from Pat
qb_pr6:AL157372        -     9.00   146.07   6.72      150660  ! AL157372 Human DNA sequence
qb_pr1:Ab055315        +     7.00   140.84   13.14     677  ! AB055315 Macaca fascicularis br
qb_in3:GGY19107        -     7.00   139.15   16.30     889  ! Y19107 Geodia cydonium mRNA for
qb_pr7:BC000537        +     7.00   135.15   27.24     1698  ! BC000537 Homo sapiens, clone I
qb_pr8:HS16P0G19       +     7.00   134.27   30.51     1959  ! 284717 Human DNA sequence from
qb_pr5:AK022627        +     7.00   133.73   32.70     2128  ! AK022627 Homo sapiens cDNA FL
qb_pr5:AK026968        +     7.00   132.98   36.01     2414  ! AK026968 Homo sapiens cDNA: FL
qb_pr10:HSU52426       +     7.00   129.79   54.17     4340  ! U52426 Homo sapiens GAK (STIM1
qb_ba1:AE005764        +     7.00   128.71   118.19    10802  ! AE005764 Caulobacter crescent
qb_pr8:HS314G4         +     7.00   117.08   276.64    21557  ! 269667 Human DNA sequence fro
qb_in2:CBRG39L12       +     7.00   115.45   340.74    41040  ! AC084573 Caenorhabditis brig
qb_htg14:AC068229      -     7.00   113.11   460.09    59928  ! AC068229 Homo sapiens chromos
qb_htg10:AC025320      +     7.00   111.79   545.12    74213  ! AC025320 Homo sapiens chromos
qb_pr8:HS434P1         +     7.00   108.57   824.27    124990  ! 297056 Human DNA sequence fr
qb_htg21:AL365221      +     7.00   108.51   830.33    126139  ! AL365221 Homo sapiens chrom
qb_htg1:AC008835       +     7.00   108.37   845.41    129045  ! AC008835 Homo sapiens chrom
qb_htg11:AC026836      -     7.00   108.18   866.26    133069  ! AC026836 Homo sapiens chrom
qb_pr6:AL162499        +     7.00   107.83   906.49    140907  ! AL162499 Human DNA sequence
qb_htg22:AL390725      +     7.00   107.80   908.91    141381  ! AL390725 Homo sapiens chrom
qb_htg15:AC072054      +     7.00   107.38   960.31    151534  ! AC072054 Homo sapiens chrom
qb_htg8:AC022965       -     7.00   107.22   979.05    155272  ! AC022965 Homo sapiens chrom
qb_htg6:AC019295       +     7.00   107.00   1.00e+03  161133  ! AC019295 Homo sapiens clone
qb_htg21:AL356004      -     7.00   106.84   1.00e+03  165366  ! AL356004 Homo sapiens chrom
qb_htg4:AC015713       -     7.00   106.83   1.00e+03  165471  ! AC015713 Homo sapiens chrom
qb_htg15:AC069216      +     7.00   106.77   1.00e+03  167251  ! AC069216 Homo sapiens chrom
qb_htg9:AC024412       +     7.00   106.37   1.10e+03  178168  ! AC024412 Homo sapiens chrom
qb_htg3:AC012334       +     7.00   106.37   1.10e+03  178453  ! AC012334 Homo sapiens clone
qb_htg15:AC069076      +     7.00   106.07   1.10e+03  187272  ! AC069076 Homo sapiens chrom
qb_htg9:AC024888       -     7.00   104.79   1.30e+03  230005  ! AC024888 Homo sapiens chrom
qb_ba2:CNSPAX04        +     7.00   103.39   1.60e+03  288440  ! AJ248286 Pyrococcus abyssi
qb_ba3:MLEPRTN3        -     7.00   102.91   1.70e+03  312050  ! AL583919 Mycobacterium leprae
qb_pat1:AX041920       +     7.00   102.20   1.70e+03  343980  ! AX041920 Sequence 815 from
qb_om:OOU297381        +     6.00   130.96   46.65     147  ! AJ297381 Oryctolagus cuniculus
qb_sts2:DM131A5S       +     6.00   130.32   50.64     163  ! Z31821 D. melanogaster STS det
qb_pr10:SOEMHDR84K     +     6.00   128.24   66.08     228  ! L32127 Saguinus oedipus MHC cl
qb_pat1:AR031837       +     6.00   128.08   67.46     234  ! AR031837 Sequence 1 from pat
qb_pat1:AR031893       +     6.00   128.08   67.46     234  ! AR031893 Sequence 1 from pat
qb_pat1:AR036274       +     6.00   128.08   67.46     234  ! AR036274 Sequence 1 from pat
qb_pat1:AR074959       +     6.00   128.08   67.46     234  ! AR074959 Sequence 1 from pat
qb_pat1:AR103823       +     6.00   128.08   67.46     234  ! AR103823 Sequence 1 from pat
qb_pr9:HSCACNLA50      +     6.00   126.22   85.61     316  ! U30995 Human cihydropyridine

seq_name: qb_pat1:AX048148

seq_fragment_plus_block:
LOCUS       AX048148     1894 bp    DNA             PAT       15-DEC-2000
DEFINITION  Sequence 1 from Patent WO0070092.
ACCESSION   AX048148
VERSION     AX048148.1  GI:11876967
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1894)
  AUTHORS   Kaufmann,J. and Xin,H.
  TITLE     Expression of ets domain proteins in cancer
  JOURNAL   Patent: WO 0070092 A 1 23-NOV 2000;
            Chiron Corporation (US)
FEATURES             Location/Qualifiers
     source          1..1894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             416..1423
                     /note="Human TEETS."
                     /codon_start=1
                     /protein_id="CAC19002.1"
                     /db_xref="GI:11876968"
                     /translation="MGSASPGLSSVSPSDLLTPFDLVSRTGLEKAAAGAVGLERDWS
                     FSPRALTEAGLSAEYLSYFDMLYFEDSSWAAKAPQASSREEPEEPEQCPVTDSQAPA
                     GSLDLVPGGLTLEEHSLFQVQSMVVGEVLKDTETACKLLNITADTMDWSPSNVQKWLL
                     WTFHQYFLPPMCFAFQELAGKELCAMSEFQLFEQQKSFLGGDVLRAAILDFWKSAAWMKER
                     ISPCATHYCASTSEESWTLSEVESSSYQGFIHLWYELKELLKEPDSYGFEIEWLNKEK
                     GILKLDSAQVAKLWSIFKNNFIAMNYDKLSKLIEQVYKKGLIEKTGTSQPLVYQFVHP
                     I"
BASE COUNT      368 a    653 c    571 g    302 t
ORIGIN


alignment_scores:
            Quality:    9.00              Length:      9
              Ratio:   1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_84_92 x AX048148  ..

 Align seq 1/1  to: AX048148  from: 1  to: 1894

       1 GluGlnProProGluGluProGluGln 9
         |||||||||||||||||||||||||||
     665 GAGGAGCCACCTGAGGAGCCTGAGCAG 691

seq_name: qb_pr4:AF071538

seq_documentation_block:
LOCUS       AF071538     1894 bp    mRNA            PRI       16-JAN-2000
DEFINITION  Homo sapiens Ets transcription factor PDEF (PDEF) mRNA, complete
            cds.
ACCESSION   AF071538
VERSION     AF071538.1  GI:4007417
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1894)
  AUTHORS   Oettgen,P., Finger,E., Sun,Z., Akbarali,Y., Thamrongsak,U.,
```

```
            Boltax,J., Grall,F., Dube,A., Weiss,A., Brown,L., Quinn,G., Kas,K.,
            Endress,G., Kunsch,C. and Libermann,T.A.
  TITLE     PDEF, a novel prostate epithelium specific ets transcription
            factor, interacts with the androgen receptor and activates
            prostate specific antigen gene expression
  JOURNAL   J. Biol. Chem. 275 (2), 1216-1225 (2000)
  MEDLINE   20092894
REFERENCE   2  (bases 1 to 1894)
  AUTHORS   Libermann,T.A., Oettgen,P., Kunsch,C. and Finger,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUN-1998) Medicine, Beth Israel Deaconess Medical
            Center, 330 Brookline Ave., Boston, MA 02215, USA
FEATURES             Location/Qualifiers
     source          1..1894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6p21.3"
                     /tissue_type="prostate"
     gene            1..1894
                     /gene="PDEF"
     CDS             416..1423
                     /gene="PDEF"
                     /note="prostate epithelium-specific"
                     /codon_start=1
                     /product="Ets transcription factor PDEF"
                     /protein_id="AAC95296.1"
                     /db_xref="GI:4007418"
                     /translation="MGSASPGLSSVSPSHLLTPPEQIVSKQGLEKAAAGAVGLERRGWS
                     PSPPATPEQGLSAFYLSYEPMLYPEQSSWAAKAPGASSREEPPPEPEQCPVIDSQAPA
                     GSLDLVPGGLTLEEHSLEQVQSMVVGEVLKDTEIACKLLNLTAEPMDWSPSNVQKWLL
                     WTEHQYRLPPMGKAFQELAGKELCAMSEEQFRQRSPEGGDVLHAHLDIWKSAAWMKER
                     TSPGATHYGASTSEESWTDSEVDSSCSGQPIHLWQFLKELLLKPHSYGRFIRWLNKEK
                     GIFKIEDSAQVAPLWGIPKNKPAMNYDKLSRSLPQYYKQGLIRKPDISQRLVYQFVHP
                     I"
BASE COUNT       368 a    654 c    571 g    302 t
ORIGIN


alignment scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_84_92 x AF071538  ..

 Align seq 1/1  to: AF071538  from: 1  to: 1894

        1 GluGluProProGluGluProGluGln 9
          |||||||||||||||||||||||||||
      665 GAGGAGCCACCTGAGGAGCCTGAGCAG 691

seq name: qb pri:AB031549

seq documentation block:
LOCUS       AB031549      1898 bp    mRNA            PRI       20-JAN-2000
DEFINITION  Homo sapiens PSE mRNA for prostate ets, complete cds.
ACCESSION   AB031549
VERSION     AB031549.1  GI:6721497
KEYWORDS    prostate ets.
SOURCE      Homo sapiens adult male prostate epithelium prostate cancer
            cell line:PC-3 cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Yamada,N., Tamai,Y., Miyamoto,H. and Nozaki,M.
  TITLE     Cloning and expression of the mouse Pse gene encoding a novel Ets
            family member
  JOURNAL   Gene 241 (2), 267-274 (2000)
  MEDLINE   20137504
REFERENCE   2  (bases 1 to 1898)
  AUTHORS   Nozaki,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-OCT-1999) to the DDBJ/EMBL/GenBank databases. Masumi
            Nozaki, Osaka University, Institute for Microbial Diseases; 3-1,
            Yamadaoka, Suita, Osaka 565-0871, Japan
            (E-mail:mnozaki@biken.osaka-u.ac.jp, Tel:81-6-879-8338,
            Fax:81-6-879-8339)
FEATURES             Location/Qualifiers
     source          1..1898
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="PC-3"
                     /cell_type="prostate cancer"
                     /chromosome="6"
                     /dev_stage="adult"
                     /map="6p21.2-21.3"
                     /sex="male"
                     /tissue_type="prostate epithelium"
                     /note="vector:pUC18"
     gene            406..1413
                     /gene="PSE"
     CDS             406..1413
                     /gene="PSE"
                     /codon_start=1
                     /product="prostate ets"
                     /protein_id="BAA89543.1"
                     /db_xref="GI:6721498"
                     /translation="MGSASPGLSSVSPSHLLTPPEQIVSKQGLEKAAAGAVGLERRGWS
                     PSPPATPEQGLSAFYLSYEPMLYPEQSSWAAKAPGASSREEPPPEPEQCPVIDSQAPA
                     GSLDLVPGGLTLEEHSLEQVQSMVVGEVLKDTEIACKLLNLTAEPMDWSPSNVQKWLL
                     WTEHQYRLPPMGKAFQELAGKELCAMSEEQFRQRSPEGGDVLHAHLDIWKSAAWMKER
                     TSPGATHYGASTSEESWTDSEVDSSCSGQPIHLWQFLKELLLKPHSYGRFIRWLNKEK
                     GIFKIEDSAQVAPLWGIPKNKPAMNYDKLSRSLPQYYKQGLIRKPDISQRLVYQFVHP
                     I"
BASE COUNT       380 a    651 c    567 g    300 t
ORIGIN


alignment scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_84_92 x AB031549  ..

 Align seq 1/1  to: AB031549  from: 1  to: 1898

        1 GluGluProProGluGluProGluGln 9
          |||||||||||||||||||||||||||
      659 GAGGAGCCACCTGAGGAGCCTGAGCAG 685

seq name: qb pat1:AX048159

seq documentation block:
LOCUS       AX048159      3317 bp    DNA             PAT       15-DEC-2000
DEFINITION  Sequence 12 from Patent WO0070092.
ACCESSION   AX048159
VERSION     AX048159.1  GI:11876975
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3317)
  AUTHORS   Kaufmann,J. and Xin,H.
  TITLE     Expression of ets-domain proteins in cancer
  JOURNAL   Patent: WO 0070092-A 12-23-NOV-2000;
            Chiron Corporation (US)
FEATURES             Location/Qualifiers
     source          1..3317
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

```
                     /note="GSEF encoding sequence with promoter (Figs. 2A 2B)"
     promoter        1. .1380
                     /note="Promoter"
     CDS             1815. .2822
                     /note "GSEF coding sequence"
                     /codon_start 1
                     /protein_id "CAC19084.1"
                     /db_xref="GI:11876976"
                     /translation "MCSASECESVEPGPPPSVEWRLSRPAAATAVLQPPPCWS
                     PSPPATPPQRLSAFYLSYFDMLYPFDSSWAAKARGASSREPPPPFPDQCPVLDSQAPA
                     GSLDLVPGGLTLEEHSLEQVQSMVVCEVLEQLETACKLLNLTALPMEWSPSNVQKWLL
                     WIEHQYRLPPMGKAFQPLAGKPLCAMSFPLPQPSPLCNVELHARLLIWESAAWMEEE
                     TSPGATHYCASTSEESWTDSEVDSSCSGQPTHLWQFLKELLLEPHSYYPFLPWLNEEE
                     GIFKIEDSAQVAKDWGIRLNNPAAMNYLRLFSLFLYYFGLIEEFLSNSLRLVYGPVHF
                     L"
BASE COUNT       710 a    1026 c     970 g     611 t
ORIGIN


alignment_scores:
            Quality:     9.00              Length:       9
              Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_84_92 x AX048159    .

 Align seq 1/1  to: AX048159  from: 1  to: 3317

       1 GluGluProProGluGluProGluGln 9
         ||||||||||||||||||||||||||||
    2064 GAGGAGCCACCTGAGGAGCCTGAGCAG 2090

seq_name: gb_pr6:AL157372

seq documentation_block:
LOCUS        AL157372    150660 bp    DNA              PRI       29-NOV-2000
DEFINITION   Human DNA sequence from clone RP11-375E1 on chromosome 6, complete
             sequence.
ACCESSION    AL157372
VERSION      AL157372.18  GI:11493239
KEYWORDS     HTG.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 150660)
  AUTHORS    Skuce,C.
  TITLE      Direct Submission
  JOURNAL    Submitted (29-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
             CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
             requests: clonerequest@sanger.ac.uk
COMMENT      On Nov 30  2000 this sequence version replaced gi:11341772
             During sequence assembly data is compared from overlapping clones.
             Where differences are found these are annotated as variations
             together with a note of the overlapping clone name. Note that the
             variation annotation may not be found in the sequence submission
             corresponding to the overlapping clone, as we submit sequences with
             only a small overlap as described above.
             This sequence has been finished according to sequence map criteria
             as follows.  An attempt is made to resolve all sequencing problems,
             such as compressions and repeats, but not necessarily within known
             annotated repeat sequence elements.  Where the sequence is
             ambiguous, there is an annotation using the 'unsure' feature key.
             The following abbreviations are used to associate primary accession
             numbers given in the feature table with their source databases:
             Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
             on the WORMPEP database can be found at
             http://www.sanger.ac.uk/Projects/C_elegans/WORMPEP/
             was generated from part of bacterial clone contigs of human
             chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
             Group.  Further information can be found at
             http://www.sanger.ac.uk/HGP/Chr6/
             RP11-375E1 is from the library RPCI-11.2 constructed at the Roswell
             Park Cancer Institute by the group of Pieter de Jong. For further
             details see: http://bacpac.chori.org/
             VECTOR: pBACe3.6
             IMPORTANT: This sequence is not the entire insert of clone
             RP11-375E1 it may be shorter because we sequence over lapping
             sections only once, except for a 100 base overlap.
             The true left end of clone RP3-391P22 is at 150561 in this sequence.
             The true right end of clone RP11-302L10 is at 113044 in this
             sequence.
FEATURES             Location/Qualifiers
     source          1. .150660
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-375E1"
                     /clone_lib="RPCI-11.2"
BASE COUNT     35470 a   38887 c   38740 g   37563 t
ORIGIN


alignment scores:
            Quality:     9.00              Length:       9
              Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2_COPY_84_92 x AL157372_2120    .

 Align seq 1/1  to reverse of: AL157372  from: 1  to: 150660

       1 GluGluProProGluGluProGluGln 9
         |||| ||||||||||||||||||||||||
  129491 GAGGAGCCACCTGAGGAGCCTGAGCAG 129465

seq_name: gb_pri:AB055415

seq_documentation_block:
LOCUS        AB055415      677 bp    mRNA            PRI       07-FEB-2001
DEFINITION   Macaca fascicularis brain cDNA, clone:qTlA-10005.
ACCESSION    AB055415
VERSION      AB055415.1  GI:12698219
KEYWORDS     fis (full insert sequence).
SOURCE       Macaca fascicularis adult male frontal lobe left cDNA to mRNA,
             clone_lib:macaque brain cDNA library qTlA clone:qTlA-10005.
  ORGANISM   Macaca fascicularis
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
             Cercopithecinae; Macaca.
REFERENCE    1  (sites)
  AUTHORS    Osada,N., Hida,M., Kusuda,J., Tanuma,R., Ikeki,K., Hirai,M.,
             Terao,K., Suzuki,Y., Sugano,S. and Hashimoto,K.
  TITLE      Isolation of full-length cDNA clones from macaque brain cDNA
             libraries
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 677)
  AUTHORS    Hashimoto,K., Osada,N., Hida,M., Kusuda,J. and Sugano,S.
  TITLE      Direct Submission
  JOURNAL    Submitted (02-FEB-2001) Katsuyuki Hashimoto, National Institute of
             Infectious Diseases, Division of Genetic Resources; 23-1, Toyama
             1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
             (E-mail:k-hashi@nih.go.jp, URL:http://www.nih.go.jp/~k-hashi/,
             Tel:81-3-5285-1111(ex.2126), Fax:81-3-5285-1181)
COMMENT      Lab host:    TOP10
             Vector:      pME18S-FL3 (Acc.No. AB009864)
             R. Site1:    BraIII (CAC!GTGIC)
             R. Site2:    BraIII (CAC!ATGIG)
             Oligo-dT primer and 5'-end cDNA was primed with an oligo(dT) primer
             [AIGTGGCCTTTTTTTTTTTTTTT]; double-stranded cDNA was synthesized
             using specific 5' and 3' primers and amplified by PCR.  The PCR
             product was digested with SfiI and size selection was performed to
             exclude fragments < 1 Kb. The SfiI-digested PCR product was cloned
```

```
            into distinct BraIII sites of pME185 FL4. XhoI sites just outside
            the BraIII sites can be used to isolate the cDNA insert.  Libraries
            were constructed by Sugano et al.(University of Tokyo, Institute of
            Medical Science). Custom primer used for sequencing
            ( 5' end primer [CTTCTGCTCTAAAAGCTGCG];
            3' end primer [CGACCTGCAGCTCGAGCACA] ).
FEATURES             Location/Qualifiers
     source          1..677
                     /organism="Macaca fascicularis"
                     /db_xref="taxon:9541"
                     /clone="QflA 10065"
                     /sex="male"
                     /tissue_type="frontal lobe left"
                     /clone_lib="macaque brain cDNA library QflA"
                     /dev_stage="adult"
BASE COUNT       298 a      85 c      99 g     195 t
ORIGIN


alignment scores:
              Quality:     7.00              Length:       7
                Ratio:    1.000               Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-9458-2-COPY_84_92 x AB055415  ..

 Align seq 1/1  to: AB055415  from: 1  to: 677

       2 GlnProProGlnGluProGlu 8
         ||||||||||||||||||||
     182 GAACCCCCAGAAGAGCCTGAA 102

Seq name: gb_in3:GCY19107

Seq documentation block:
LOCUS        GCY19107       889 bp     mRNA             INV       11-JUL-2000
DEFINITION   Geodia cydonium mRNA for GCDD2 protein (GCDD2 gene).
ACCESSION    Y19107
VERSION      Y19107.1  GI:9187967
KEYWORDS     GCDD2 gene; GCDD2 protein.
SOURCE       Geodia cydonium.
  ORGANISM   Geodia cydonium
             Eukaryota; Metazoa; Porifera; Demospongiae; Tetractinomorpha;
             Astrophorida; Geodiidae; Geodia.
REFERENCE    1  (bases 1 to 889)
  AUTHORS    Wiens,M., Krasko,A., Muller,C.I. and Muller,W.E.
  TITLE      Molecular evolution of apoptotic pathways: cloning of key domains
             from sponges (bcl-2 homology domains and death domains) and their
             phylogenetic relationships
  JOURNAL    J. Mol. Evol. 50 (6), 520-531 (2000)
  MEDLINE    20296813
REFERENCE    2  (bases 1 to 889)
  AUTHORS    Mueller,W.E.G.
  TITLE      Direct Submission
  JOURNAL    Submitted (22-JUN 1999) W.E.G. Mueller, Abteilung fuer Angewandte
             Mol. Biologie, Institut fuer Physiologische Chemie, Universitaet,
             55099 Mainz, GERMANY
FEATURES             Location/Qualifiers
     source          1..889
                     /organism="Geodia cydonium"
                     /db_xref="taxon:6047"
     gene            107..799
                     /gene="GCDD2"
     CDS             107..799
                     /gene="GCDD2"
                     /codon_start=1
                     /protein_id="CAB97128.1"
                     /db_xref="GI:9187968"
                     /translation="MAEQDAEQVVEEAPEPGIVIIDHHTPDEPVAEPVQEVSLAPA
                     SDGDIDDIIDRDGFDGPDPEVRRELVNEWAVENGSQATYVAEVLALEAARMEVAEALEVCNL
                     LKGQDGATPGSMATDGSSGGSSGSAPDGRLGDMPELGELEDSSKEDDVEAEPDTDVEE
                     EGTKTLEEENGAEVEVIEEEPGTDAEKEEPGAEEAGLEGPTEPGIVIWGTLVEVEEEEE
                     KELTKKIREALQ"
BASE COUNT       272 a      175 c      259 g      182 t        1 others
ORIGIN


alignment scores:
              Quality:     7.00              Length:       7
                Ratio:    1.000               Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-9458-2-COPY_84_92 x GCY19107/rev  ..

 Align seq 1/1  to reverse of: GCY19107  from: 1  to: 889

       1 GlnGluProProGlnGluPro 7
         |||||||||||||||||||||
     477 GAGGAACCTCCAGAGGAACCT 457

Seq name: gb_pri:BC000567

Seq documentation block:
LOCUS        BC000567      1698 bp    mRNA             PRI       16-MAR-2001
DEFINITION   Homo sapiens, clone IMAGE:4162384, mRNA, partial cds.
ACCESSION    BC000567
VERSION      BC000567.1  GI:12654540
KEYWORDS     .
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 1698)
  AUTHORS    Strausberg,R.
  TITLE      Direct Submission
  JOURNAL    Submitted (15-NOV-2000) National Institutes of Health, Mammalian
             Gene Collection (MGC), Cancer Genomics Office, National Cancer
             Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
             USA
  REMARK     NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT      Contact: Robert Strausberg, Ph.D.
             Tel: (301) 496-1550
             Email: Robert_Strausberg@nih.gov
             Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: National Institutes of Health Intramural
             Sequencing Center (NISC),
             Gaithersburg, Maryland:
             Web site:          http://www.nisc.nih.gov/
             Contact:           nisc_mgc@nhgri.nih.gov
             Shevchenko,Y., Wetherby,K.D., Beckstrom-Sternberg,S.M.,
             Benjamin,R., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
             Dietrich,N.L., Guan,X., Gupta,J., Ho,S. L., Karlins,E., Legaspi,R.,
             Lim,M., Maduro,Q.L., Masiello,C., Mastrian,S.D., McCloskey,J.C.,
             Mebwell,J., Pearson,R., Snyder,B., Stantripop,S., Thomas,P.J.,
             Tomlson,E.E., Touchman,J.W., Tsurgeon,C., Voqt,J.L., Walker,M.A.,
             Zhang,L.-H. and Green,E.D.

             Clone distribution: MGC clone distribution information can be found
             through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
             Series: IRAL Plate: 5 Row: a Column: 5.
FEATURES             Location/Qualifiers
     source          1..1698
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4162384"
                     /tissue_type="Brain, neuroblastoma"
                     /clone_lib="NIH_MGC_19"
                     /lab_host="DH10B R"
                     /note="Vector: pCR2"
     CDS             <1..1573
                     /codon_start=2
                     /product="Unknown (protein for IMAGE:4162384)"
```

```
                     /protein_id="AAH00537.1"
                     /db_xref="GI:12653541"
                     /translation="ARGLPVLLLLLRASCPWCQPQCARSPSFPPPEEEIPEEDGLLV
                     LSRHTLGLALRPHPALLVEEYAPWCGHCQALAPEYSKAAAVLAAESMVVILAKVDGPA
                     QRELAEEFGVTEYPTLKFFRNGNRTHPEEYTGPRDAEGIAEWLRKRVGPSAMRLETEA
                     AAQALIGGRDLVVIGFFQDLQDEDVATFLALAQDALDMTFGLTDRPRLFQQPGLTKDT
                     VVLFKKFDEGRADFPVDPETLGLDLGDLSKFLVTHSMPLVTEFNSQTSAKIFAAPILNH
                     LLLFVNQTLAAHRELLAGFGFAAPRFRGQVLFVVVDVAADNFHVLQYFGLKAEAAPTL
                     RLVNLETTKKYAPVEGGPVTAASITAFCHAVLNGQVKPYLLSQELPPDWDQRPVKLLV
                     GKNFEQVAFDETKNVFVEFYAPWCTHCKEMAPAWEALAEKYQHEDLLIAEMDATANE
                     LDAFAVHGFPTLKYFPAGPDRRVIEYEGTRTLEDFINFIEENRTVLPTEEPIEEPAAPT
                     PEPPANSTMGSKEEL"
BASE COUNT       339 a    502 c    543 g    414 t
ORIGIN


alignment_scores:
            Quality:     7.00            Length:       7
              Ratio:    1.000              Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_84_92 x BC000537     .

 Align seq 1/1  to: BC000537  from: 1  to: 1998

       1 GluGluProProGluGluPro 7
         |||||||||||||||||||||
    1493 GAGGAGCCCCCGGAGGAGCCA 1513

seq_name: gb_pr8:HS16PQG19

seq_documentation_block:
LOCUS       HS16PQG19     1959 bp    mRNA            PRI       04 FEB 1997
DEFINITION  Human DNA sequence from cDNA 16pHq6;19 from chromosome 16pl3.3.
ACCESSION   Z84717
VERSION     Z84717.1  GI:1817571
KEYWORDS    16pl3.3; cDNA.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1959)
  AUTHORS   Flint Higgs,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (30 JAN 1997) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1RQ, UK. E-mail enquires: humpub@sanger.ac.uk
COMMENT     This clone was sequenced at the Institute of Molecular Medicine,
            Oxford. This sequence contains simple reads. An attempt is made to
            resolve sequencing problems, such as compressions and repeats, but
            some sequence is ambiguous.
FEATURES             Location/Qualifiers
     source          1..1959
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16pl3.3"
                     /clone="16pHq6.19"
BASE COUNT       347 a    574 c    664 g    372 t      2 others
ORIGIN


alignment_scores:
            Quality:     7.00            Length:       7
              Ratio:    1.000              Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_84_92 x HS16PQG19    ..

 Align seq 1/1  to: HS16PQG19  from: 1  to: 1959

       1 GluGluProProGluGluPro 7
         ||||||||||||||||||||||
    1778 GAGGAGCCCCCGGAGGAGCCA 1798

seq_name: gb_pr5:AK022627

seq_documentation_block:
LOCUS       AK022627     2138 bp    mRNA            PRI       29-SEP-2000
DEFINITION  Homo sapiens cDNA FLJ12565 fis, clone NT2RM4000848.
ACCESSION   AK022627
VERSION     AK022627.1  GI:10434126
KEYWORDS    ...lete sapiens, fis (full insert sequence).
SOURCE      Homo sapiens teratocarcinoma cell line:NT2 cDNA to mRNA.
            clone_lib:NT2RM4 clone:NT2RM4000848.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
            Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
            Watatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
            Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
            Imai,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
            Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
            Nakamura,Y., Nagahari,K., Masuho,Y., Nimomiya,K. and Iwayanagi,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished (2000)
REFERENCE   2  (bases 1 to 2138)
  AUTHORS   Isogai,T. and Otsuki,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (23 AUG 2000) to the EMBL/GenBank/DDBJ databases. Takao
            Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,
            Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing; Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing and clone selection;
            Helix Research Institute (supported by Japan Key Technology Center
            etc.) and Department of Virology, Institute of Medical Science,
            University of Tokyo.
FEATURES             Location/Qualifiers
     source          1..2138
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="NT2"
                     /cell_type="teratocarcinoma"
                     /clone="NT2RM4000848"
                     /clone_lib="NT2RM4"
                     /note="cloning vector: pME18SFL3 mRNA from uninduced NT2
                     neuronal precursor cells."
     CDS             47..1915
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAB14109.1"
                     /db_xref="GI:10434127"
                     /translation="MEIEEPLSISLEVEVPTLSVPGPIPIIIPCSPSWNPSLIIGKFPE
                     EAIQAFLELNSSLEVLIGAVMMYNLSVNQQTIKMVGVSVDVVNEYAMALKCLEDKLKPCF
                     KEKKDIIAELIKSQKVFSEKLDHLSKELAWVHATVSSQLKMLIIYWLLKVCLRIIEEG
                     DHIGSLFAFMPEFYLSVATNSYSALKNYFGPVHSMEELPQYEETLIRLAATLAKHEAD
                     ARIVGIDIKGSLMQAIASYVCYPHSLRAVEKIPEEQKTAMVRNLLAPYPQKLWAQTNW
                     ILVRLWRQGPGYRYTRLPHLLKTKLEDANLPSLQKFCPSILLQQHMAGLLQQGPLVA
                     PSFLNSVLNQLLNWAFSEFIGMIQEIQQAAEKLERNFVDSKQLKVCAICFDLSVSLLKV
                     LEMIIILVPEIFLDWIRPISEMLLRRLAQLLNQVLNRVIAERNLEFDKVVLRFIGLES
                     VIDHYPILVAVIGILVQLLVRGPASEREQAISVLLADPCFQLRSICYELGQPEPPAPGI
                     ALPAPDKRKPFSLQSYADYISADELAQVFQMLAHLTSASAQAAAASLPLSEEDLCPIPY
                     AHPISAAFQLPVHFSCFACNQHLMNNEFYFFCKLLTVSVDWFFGANISFISSAA"
BASE COUNT       436 a    683 c    581 g    438 t
ORIGIN


alignment_scores:
            Quality:     7.00            Length:       7
              Ratio:    1.000              Gaps:       0
```

```
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-126-945B-2_COPY_84_92 x AK022627   .

 Align seq 1/1   to: AK022627   from: 1   to: 2138

        3 ProProGluGluProGluGln 9
          ||||||||||||||||||||
      170 CCCCCCGAGGAGCCTGAGCAG 190

Seq name: gb pr%:AK026968

Seq documentation block:
LOCUS       AK026968      2414 bp    mRNA            PRI       29-SEP-2000
DEFINITION  Homo sapiens cDNA: FLJ23415 fis, clone HEP12021.
ACCESSION   AK026968
VERSION     AK026968.1  GI:10439956
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens hepatoma cell line:HepG2 cDNA to mRNA, clone lib:HEP
            clone:HEP12021.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Kawabata,A., Hikiji,T., Kobatake,N., Inuzuka,H., Ikoma,Y.,
            Okamoto,S., Okitani,R., Ota,F., Suzuki,Y., Obayashi,M., Nishi,T.,
            Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished (2000)
REFERENCE   2  (bases 1 to 2414)
  AUTHORS   Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
            Shibahara,T., Tanaka,T. and Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumi
            Sugano, Institute of Medical Science, University of Tokyo,
            Laboratory of Genome Structure Analysis, Human Genome Center,
            Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
            (E-mail:ednal@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
            Fax:81-3-5449-5416)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5' & 3' end one pass sequencing: Department of
            Virology and Human Genome Center, Institute of Medical Science,
            University of Tokyo (partly supported by Science and Technology
            Agency).
FEATURES             Location/Qualifiers
     source          1..2414
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="Hep-G2"
                     /cell_type="hepatoma"
                     /clone="HEP12021"
                     /clone_lib="HEP"
                     /note="cloning vector pME18SFL3"
     CDS             41..2349
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAB15697.1"
                     /db_xref="GI:10439957"
                     /translation="MEMEGEEYMVHLQPQEVALDVWQIYEQQTGVAGQLQRLLEVEL
                     PGVEYNGKVMDKHFILKGEFPAGQEYLVGSGMEANLASLQAWVKQAEGEYVRDAVIPI
                     PAPAMAQRPMQALAVGAELLELPQLGWLSSVFLGKANKELKDAAVCWLLKSWKLCSKE
                     VKVECLCVDQEREWTWNSSGVLTEAGVPNMYNLVGMIDGKMVESDVNEYAMALMQTLE
                     RLAWVHATVYGQLKMLTDYWLLVSKEAIHIPKDEEGSWIVGAKMPHRYTVSAVNVLCA
                     NYFGPVHSMKELPGYREELLTRLAALLAKHPADARLVCDLVDSLMQALASYVCYHRSL
                     RAVERTPEDQRLAMVRNLLAPYLGKFWAQLNWCLVKLWCPVGTGYKYLKLFHELLEKL
                     EDANLPSLQKFCTPSLLGQGHMADLLQQCPDVAPSFLNCVLNQLNWAFSIGFTGMLQE
                     QAAERLERNFVDSRQLKVCATPFDLSVSLLRVLEMITTLVPELPLDWFRPTSEMLLRR
                     LAQLLNLVLNRVIAENNLFPRVVTLRLPSKLVHIYPLLVAVLSLLVQLLVRGPACSER
                     EQATSVLLADPCFQLPSLQYLLQQPPAPCLALAPGPPPPPSLQSYAPVLSAPPLAL
```

```
                     VEQMLAHLLSASAQAAAASLPTSEELCPLCYAHPLSAVFQPSGHKSCKACLNQHLMN
                     NKLCPFCKLITSNLLACLYPHWWEPSHGPNCA"
BASE COUNT       501 a     769 c     657 g     487 t
ORIGIN


Alignment scores:
                Quality:     7.00              Length:       7
                  Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-126-945B-2_COPY_84_92 x AK026968   ..

 Align seq 1/1   to: AK026968   from: 1   to: 2414

        3 ProProGluGluProGluGln 9
          ||||:|||||||||||||||||
      938 CCCCCCGAGGAGCCTGAGCAG 918

Seq name: gb pri:BS05426

Seq documentation block:
LOCUS       BS05426      4040 bp    mRNA            PRI       18-JUL-1997
DEFINITION  Homo sapiens G0K (STIM1) mRNA, complete cds.
ACCESSION   S05426
VERSION     S05426.1  GI:226434
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4040)
  AUTHORS   Parker,N.J., Begley,C.G., Smith,P.J. and Fox,R.M.
  TITLE     Molecular cloning of a novel human gene (D11S4896E) at chromosomal
            region 11p15.5
  JOURNAL   Genomics 37 (2), 253-256 (1996)
  MEDLINE   97079692
REFERENCE   2  (bases 1 to 4040)
  AUTHORS   Parker,N.J., Begley,C.G., Smith,P.J. and Fox,R.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAR-1996) Nigel J. Parker, Haematology/Oncology,
            Royal Children's Hospital, Flemington Rd., Parkville, VIC 3052,
            Australia
COMMENT     On Jul 18, 1997 this sequence version replaced gi:1322051.
FEATURES             Location/Qualifiers
     source          1..4040
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11p15.5"
                     /note="located immediately telomeric of FRM1 gene; DNA
                     number: D11S4896E"
     gene            1..4040
                     /gene="STIM1"
     CDS             646..2623
                     /gene="STIM1"
                     /codon_start=1
                     /product="G0K"
                     /protein_id="AAC51627.1"
                     /db_xref="GI:226435"
                     /translation="MKVVECACWCLWLCCLHLGCQCSCRLRCRRCCSSCVANSLCST
                     AALEPCFLDKFLCTSEDEKLSEEAVRNTHKLMDDDANGDVDVEEFSDEFLKEDLNYHD
                     VEGCLTKDGLELTVLPLWLEAWEEKCVYWWIVCVVGWILLYVFFLPCYIFIPFLGS
                     GHAMGHLAAVINPLMKYVLKMIIECTHECHEGKQLKAIDCVLPCPPLLTLIHNHLKDEM
                     LVVCVLCVGWLAYSQNKYSKEHMEMMFLLCGHPAEQSLHFLQDPLPAQEERRIVE
                     VLRVHLLLKELLDLNLAECQAGFLPLLLKFCGLNLPSSPQRYADLLLTLCVREALPKAEKE
                     LESHSSWYAPEALQKWLQLTHEVEVCQYYNLKKQNAEKQLLVAKFQAEKIKKKKNILFG
                     LTHVAHSCSLLCVCWKLLCAKQALSEVLAALFLKGLHRWQGLLLLCGLQLVNNPGLHSL
                     VAALNLDPSWMGSTRPNPAHFLMTDDVDDMDEELVSPLSMQSPSLQSSVRQRLTEPQH
                     GLGSQRDLTHSDSESSLHMSDRQRVADKPPQMSKAADEALNAMLSNGHRELLEGVHPG
                     CVEVEFLFLSPALAKKALLALNLCLDKAHCLRFLSPCAPPSSPPVCSCMPFGSFDP
                     CHLPVGLSPALQASPNLPLPHLAGPFAVAEFDNSLDFLLSSSLSPEEFGLFEEP
```

```
                     LKK"
     variation       1645
                     /gene "STIM1"
                     /replace "q"
BASE COUNT      893 a   1149 c   1114 g    884 t
ORIGIN


alignment_scores:
            Quality:     7.00            Length:       7
              Ratio:    1.000              Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_84_92 x HS052126/rev    1..

 Align seq 1/1  to reverse of: HS052126  from: 1  to: 1943

        1 GluGluProProGluGluPro 7
          |||||||||||||||||||||
      614 GAGGAGCCCCCAGAGGAGCCA 593

seq_name: gb_bal:AE005764

seq_documentation_block:
LOCUS       AE005764     10802 bp     DNA             BCT       28-MAR-2001
DEFINITION  Caulobacter crescentus  section 93 of 95 of the complete genome.
ACCESSION   AE005764 AE005673
VERSION     AE005764.1  GI:13422130
KEYWORDS    .
SOURCE      Caulobacter crescentus.
  ORGANISM  Caulobacter crescentus
            Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
            Caulobacter.
REFERENCE   1  (bases 1 to 10802)
  AUTHORS   Nierman,W.C., Feldblyum,T.V., Laub,M.T., Paulsen,I.T., Nelson,K.E.,
            Eisen,J., Heidelberg,J.F., Alley,M.R.K., Ohta,N., Maddock,J.R.,
            Potocka,I., Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D.,
            Ely,B., DeBoy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L.,
            Haft,D.H., Kolonay,J.F., Smit,J., Craven,M.B., Khouri,H.,
            Berry,K., Utterback,T., Tran,K., Wolf,A., Vamathevan,J.,
            Ermolaeva,M., White,O., Salzberg,S.L., Venter,J.C., Shapiro,L. and
            Fraser,C.M.
  TITLE     Complete genome sequence of Caulobacter crescentus
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)
  MEDLINE   21173698
REFERENCE   2  (bases 1 to 10802)
  AUTHORS   Nierman,W.C., Feldblyum,T.V., Paulsen,I.T., Nelson,K.E., Eisen,J.,
            Heidelberg,J.F., Alley,M.R.K., Ohta,N., Maddock,J.R., Potocka,I.,
            Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D., Ely,B.,
            Laub,M.T., DeBoy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L.,
            Haft,D.H., Kolonay,J.F., Smit,J., Craven,M.B., Khouri,H., Shetty,J.,
            Berry,K., Utterback,T., Tran,K., Wolf,A., Vamathevan,J.,
            Ermolaeva,M., White,O., Salzberg,S.L., Shapiro,L., Venter,J.C. and
            Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1..10802
                     /organism="Caulobacter crescentus"
                     /db_xref="taxon:69394"
     gene            complement(86..262)
                     /gene "CC0870"
     CDS             complement(86..262)
                     /gene "CC0870"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAK22855.1"
                     /db_xref="GI:13422131"
                     /translation="MAFKFRAAAFGALMLVAAGAAQAAVTPAPVQQPLAPFDQIKPAS
                     PAPDGSVWVKILLW"
     gene            complement(341..925)
                     /gene "CC0871"
     CDS             complement(341..925)
                     /gene "CC0871"
                     /note "identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAK22856.1"
                     /db_xref="GI:13422132"
                     /translation="MLLCRAAASLPFTFCWNCFMFLLLKRFGLA[illegible]VNGAYRG
                     ELAAQGWLSESFLLGGQKIDFASLRADLAAKPGSLLLTFRGDLAWFKLFAVAWIFQAD
                     DVWYVGMLTVTPFRQAGGMCRAVLETCEAYARARGCKVMRMTVISVELALTAWYERRG
                     YKLTGELYGFLVGDARGLRQKIDLARVVMLKGL"
     gene            964..1761
                     /gene "CC0872"
     CDS             964..1761
                     /gene "CC0872"
                     /note "identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAK22857.1"
                     /db_xref="GI:13422133"
                     /translation="MPDLIGSRPMVRAMQARIVSEYRHFVKCFTPEEIDRAGLAAGACF
                     PEDGLYAVELVESGPFAAPARTSAMRLIVLARIAVARAMIAYIPKSVLIAKAFGF
                     PDFAGDLRQASGFAGFFAWLALDLGDMAQCPLKVIESGSARKFMDSKSCYVSIVNLAS
                     VRDLSAKVSREFLSFLRFRANLYVEGWFAWAINDLEDQLFLSAGAKAAVLKFIVRCAAI
                     HVDPISAERDIDLVKALFDHHGHMFCGIYLNVIEGGAVEKGDVVAAG"
     gene            complement(1762..2052)
                     /gene "CC0873"
     CDS             complement(1762..2052)
                     /gene "CC0873"
                     /note "identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAK22858.1"
                     /db_xref="GI:13422134"
                     /translation="MFRYFLSFPAFAIIGGIWTYSPGRWCVFGAAIVAKFLGAIIEMI
                     AEHPGMGGPDENLRAGYRRCASGSHVVFYKVGVKVEIIRVLHQSMNAKAHIG"
     gene            complement(2049..2315)
                     /gene "CC0874"
     CDS             complement(2049..2315)
                     /gene "CC0874"
                     /note "identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAK22859.1"
                     /db_xref="GI:13422135"
                     /translation="MAAKNLGVVLGmrqALtGqVAGGrGSA[illegible]KAAIRLIEEN
                     EAKLAALRAALIEGEESGFIEDFDFDAFTEERSRASAPQGFHEE"
     gene            2305..3258
                     /gene "CC0875"
     CDS             2305..3258
                     /gene "CC0875"
                     /note "identified by match to  protein family hbmK"
                     /codon_start=1
                     /transl_table=11
                     /product="hbmK family protein"
                     /protein_id="AAK22860.1"
                     /db_xref="GI:13422136"
                     /translation="MLAMADLYQNLVPHSISPPCPPPAKAPPMEIIVKAWIAAKERL
                     KDAGIDQPSIDARLMLEVAAGVTRTEIVTDPYRELSAEQIAILNDVIERRAKREPVSH
                     IIGREGFWELILQVNENVLIPPPETEVIVDEVLEAETEHMAFSMELIVVPVILLLAV
                     LAERPAAKGLGIDASSEALAVAKENAANLDLNIRAALLDGIWTIGISDSIDLVVSNP
                     PYIPIEVIDTLEPEVRIHEPRLALDGGPDGLAAYREIAPELLRVLKPGGLFAVEIGYD
                     QSQAVEALFSAAGATLVKIVKDLSIHDRVVIGVKNPLESPA"
     gene            3517..4629
                     /gene "CC0876"
     CDS             3517..4629
```

```
                     /gene "CC0876"
                     /note "identified by Glimmer2; putative"
                     /codon_start 1
                     /transl_table 11
                     /product "hypothetical protein"
                     /protein_id "AAK22861.1"
                     /db_xref "GI:13422147"
                     /translation "MKDEKGMKEQEKRNNEQGNQGKFQQHNANEAEQNQFEAVKVKG
                     AAQSVYEKYQQLARDASSSGDRVLAENYIQHAEHYFRVIKATQPNRPVADIVGKDVYA
                     AYEIDFEAEPPEEPEVVEQPAEQAEGGEGGENGEPRRDREENRDRNRDRDRNEENRD
                     RDRNRDDRNRDGDSKRDRWRGEDDRPRDDRPRDDRPRDERPRDERPRDDRPRDERPRE
                     DRERDRDDRPRDDRPRDDRPRDDRPAAEGEQPQGEGEFRESKRRRDRNRDRDRDWQPRGERDP
                     MAVIEPEASPLTAEAPAGEGSEMLKQVDGAVSHAPAFLGKAEKAPEAASSAPAAEAA
                     PASDDGEEVEKPRKKAPEQFEAASSATAESSGES"
gene                 4797..5552
                     /gene "CC0877"
CDS                  4797..5552
                     /gene "CC0877"
                     /note "identified by Glimmer2; putative"
                     /codon_start 1
                     /transl_table 11
                     /product "conserved hypothetical protein"
                     /protein_id "AAK22862.1"
                     /db_xref "GI:13422148"
                     /translation "MTLPRCIIPLLIAAACALGGCAYNPTGLGRDGLLIVNDNNLARDG
                     EKAWADALRTSHVSKEPKKNAKVQAVGQRVIVAAGLAGRPWEYAVFLEEAPNAFVLPG
                     GYVGVTVGLLSLVENDDQLAAVIGHEAGHVVARHAAERVSQETISKVILGTAKAAAGG
                     TEFGKLLEEHGEQATRYGLLGFSPEQEQRLEAQEIGVELMQRAGYDPEELAVRIMKNMQA
                     LDGGKSGGDLASTHPSDAVRIQELERYLASKGW"
gene                 5651..8230
                     /gene "CC0878"
CDS                  5651..8230
                     /gene "CC0878"
                     /note "identified by match to  protein family HMM"
                     /codon_start 1
                     /transl_table 11
                     /product "ATP-dependent Clp protease, ATP-binding subunit
                     ClpB"
                     /protein_id "AAK22863.1"
                     /db_xref "GI:13422149"
                     /translation "MNIDQVSQEAKQAVQSAQSLATAEQQQFADTEHLLKVLLEDKE
                     LSRALIQSAGGDPDQLDGGVETLLAKTPKVSGAGQLYMEPTAKVFALAEKEAKAAG
                     DAFVTTERLLIAIAKEGGEAAKLFKEAGVSAQSLEIAANAMKKGKTADSANAEEGYEA
                     LKKYARDLTAAARDGKLDPVIGRDEEIRRTIQVLSRRTKNNPVLIGEPGVGKTAIVEG
                     LALRIVNGDVPESLKDKKLLSLDMGSLIAGAKYRGEFEERLKAVLKEVIAAEGSILF
                     IDEMHTLVGAGKGDGAMDASNLLKPALARGELHCVGATTLDEYRKHVEKDAALARRFQ
                     PVFVSEPTVEDTVSILRGLKEKYEVHHGVRISDSAIVAAATLSNRYIADRFLPDKAID
                     LVDEASSRVRMQIDSKPEELDELDRKIVQLKIEKEALSKEDAASKQRLENLEVELSD
                     LQERSDEMTARWKAEKEKVQGAAQAREALDRLRADLANAQRAGDFARAGQLQYGEIPA
                     LERRLAEAEAGDTQALTPEVVDAEQIAAVVSRWTGVPVEKMLEGERERKLLKMEDELHG
                     RVVGQDEALEAVSDAVRRARAGLQDPSKPIGSFLFLGPTGVGKTELTKSLAEFLFADE
                     AALTRMDMSEYMEKHSVSRLIGAPPGYVGYDEGGALTEAIRRRPYQVVLFDEIEKAHP
                     DVFNVLLQVLDDGRLTDGQGRTVDFRNTLIIMTSNLGAEYLASQEDGEDVEAVRPMVM
                     NTVRGHFRPEFLNRIDETILFKRLSRHNMGDIVRIQLQRVEKLLADRRMALALDAEAL
                     NWLADKGYDPVYGARPLKRVIQKELVDPLAKELLAGELHDGGVIAVGVIDKQLSFGKA
                     VLQ"
gene                 8241..8782
                     /gene "CC0879"
CDS                  8241..8782
                     /gene "CC0879"
                     /note "identified by Glimmer2; putative"
                     /codon_start 1
                     /transl_table 11
                     /product "hypothetical protein"
                     /protein_id "AAK22864.1"
                     /db_xref "GI:13422150"
                     /translation "MLVSLCFVVSWFFEFVGPQLFFICAALLGAIFSSCVFI
                     SSVAFSDFSREGRPGRAFSFFPVALSFFMAKTFKLIFSPQTFVVGDPYVWFLHFLPQ
                     WSYDFSEMHGKVIAAPHFFQFFAGLFFFSFFFFSAFFLAFSSLFVRFVASAIASA
                     IFAGWFPLSRGKFVVIYVDEEGA"
gene                 complement(8792..9070)
                     /gene "CC0880"
CDS                  complement(8792..9070)
                     /gene "CC0880"
                     /note "identified by Glimmer2; putative"
                     /codon_start 1
                     /transl_table 11
                     /product "conserved hypothetical protein"
                     /protein_id "AAK22865.1"

alignment scores:
                Quality:     7.00              Length:       7
                  Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 03 09 126 945b 2 copy 84 92 x AE005764  ..

 Align set 1/1  to: AE005764  from: 1  to: 10802

        2 GlimmerProductionProtein 8
          |||| ||||||||||||||||||
     9847 GAGCGCCGGAAGAGCCCGAA 9867

seq name: gb pr8:HS414G4

seq documentation block:
LOCUS       HS414G4        41557 bp    DNA             PRI       23-NOV-1999
DEFINITION  Human DNA sequence from cosmid 414G4 from a contig from the tip of
            the short arm of chromosome 16, spanning 2Mb of 16p13.3. Contains
            ESTs and CpG islands.
ACCESSION   Z99667
VERSION     Z99667.1  GI:1200412
KEYWORDS    16p13.3; CpG island.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 41557)
  AUTHORS   Thomas,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-1996) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1RQ, England. E-mail enquiries regarding this sequence:
            humquery@sanger.ac.uk
COMMENT     IMPORTANT:
            This sequence is not the entire insert of clone 414G4. It may be
            shorter because we only sequence overlapping sections once, or
            longer because we arrange for a small overlap between neighbouring
            submissions.
            The true right end of clone 414G4 is at 41557.
            414G4 is from a 280kb clone contig extending from the telomere of
            16p. Higgs D.R., Flint J. unpublished. MRC Molecular Haematology
            Unit, Institute of Molecular Medicine, Oxford.
            414G4 came from the Los Alamos, flow sorted human chromosome 16
            libraries constructed by Norman Doggett (unpublished)
FEATURES             Location/Qualifiers
     source          1..41557
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16p13.3"
                     /clone="LA16_414G4"
                     /clone_lib="LA16"
     repeat_region   428..716
                     /note="Alu repeat; matches 308..1 of consensus"
     misc_feature    914..1691
                     /note="expressed region  matched by multiple ESTs"
     misc_feature    1692..2316
                     /partial
                     /note="match: 5' EST 351457 clone 91798"
     misc_feature    2290..2810
                     /partial
                     /note="match: 5' EST 187064 clone 115415"
     misc_feature    complement(3171..3720)
                     /note="expressed region  matched by multiple ESTs"
     repeat_region   5136..5489
```

```
                     /note="3 copies of 118 mer 88 % conserved"
     misc_feature    5908. .6086
                     /note="match: EST Z39463 clone c-lqq06"
     repeat_region   7536. .7567
                     /partial
                     /note="Alu repeat: matches 1. .32 of consensus"
     repeat_region   7568. .7856
                     /note="Alu repeat: matches 1. .308 of consensus"
     repeat_region   9446. .9744
                     /note="Alu repeat: matches 308. .1 of consensus"
     misc_feature    12398. .13777
                     /note="putative CpG island"
     repeat_region   13241. .13275
                     /note="7 copies of 5 mer 86 % conserved"
     repeat_region   14561. .14677
                     /partial
                     /note "Alu repeat: matches 254. .136 of consensus"
     repeat_region   14680. .14971
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   14974. .15106
                     /partial
                     /note="Alu repeat: matches 145. .1 of consensus"
     repeat_region   15127. .15231
                     /partial
                     /note "Alu repeat: matches 126. .1 of consensus"
     repeat_region   15248. .15535
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   15607. .15917
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   16005. .16114
                     /partial
                     /note="Alu repeat: matches 108. .1 of consensus"
     repeat_region   16123. .16283
                     /partial
                     /note="Alu repeat: matches 309. .150 of consensus"
     misc_feature    17661. .18594
                     /note="putative CpG island"
     repeat_region   18912. .19019
                     /note="3 copies of 36 mer 89 % conserved"
     repeat_region   19514. .19621
                     /note="MER22 element fragment"
     misc_feature    19922. .20499
                     /note="match: 5' EST H21176 clone 173892"
     repeat_region   20504. .20539
                     /note="4 copies of 9 mer 97 % conserved"
     repeat_region   21472. .21766
                     /note="Alu repeat: matches 308. .1 of consensus"
     misc_feature    25040. .25378
                     /note="expressed region - matched by multiple ESTs"
     misc_feature    complement(25620. .25910)
                     /note "expressed region - matched by multiple ESTs"
     repeat_region   26707. .26817
                     /partial
                     /note="Alu repeat: matches 298. .184 of consensus"
     repeat_region   26818. .26963
                     /partial
                     /note="Alu repeat: matches 160. .1 of consensus"
     misc_feature    complement(27039. .27209)
                     /note="expressed region - matched by multiple ESTs"
     repeat_region   30474. .30611
                     /note="3 copies of 46 mer 94 % conserved"
     misc_feature    31087. .31220
                     /note="expressed region - matched by multiple ESTs"
     repeat_region   31496. .31557
                     /partial
                     /note="Alu repeat: matches 1. .62 of consensus"
BASE COUNT      5713 a   9969 c   9918 g   5957 t
ORIGIN


alignment_scores:
            Quality:    7.00            Length:       7
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_84_92 x BS414G4    ..

 Align seq 1/1  to: BS414G4  from: 1  to: 41557

        1 GluGluProProGluGluPro 7
          |||||||||||||||||||||
   24510 GAGGAGCCCCAGGAGGAGCCA 24530

Seq_name: gb_in2:CBRG39L12

Seq_document_file_block:
LOCUS       CBRG39L12   41040 bp    DNA             INV       04-NOV-2000
DEFINITION  Caenorhabditis briggsae cosmid G39L12, complete sequence.
ACCESSION   AC084573
VERSION     AC084573.1  GI:11095023
KEYWORDS    HTG.
SOURCE      Caenorhabditis briggsae.
  ORGANISM  Caenorhabditis briggsae
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 41040)
  AUTHORS   Washington University Genome Sequencing Center.
  TITLE     The C. briggsae Genome Sequencing Project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 41040)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     Submitted by:
                  Genome Sequencing Center
                  Department of Genetics, Washington University.
                  St. Louis, MO 63110, USA
                  e-mail: jspieth@watson.wustl.edu


            NOTICE:  This sequence may not be the entire insert of this clone
            as may be obtained because we edit, screen, or collapse vectors
            once, or longer because we provide a small overlap between
            neighboring submissions.
FEATURES              Location/Qualifiers
     source           1. .41040
                      /organism="Caenorhabditis briggsae"
                      /strain "GuJArat G16"
                      /db_xref="taxon:6238"
                      /clone "C39L12"
BASE COUNT     12385 a   8156 c   7610 g   12889 t
ORIGIN


alignment_scores:
            Quality:    7.00            Length:       7
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_84_92 x CBRG39L12    ..

 Align seq 1/1  to: CBRG39L12  from: 1  to: 41040

        1 GluGluProProGluGluPro 7
          |||| ||||||||||||||||
     6842 GAAGAACCTCCAGAAGAACCT 6862
```

```
OM of: US-09-126-945B-2_COPY_84_92 [illegible]

Date: Sep 14, 2001 9:48 PM

About: Results were produced by the [illegible]
       Copyright (c) [illegible] Compugen Ltd

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=[illegible]/US09126945/[illegible]/app_query.fasta_1.1067
 -DB=N_Geneseq_0601 -QFMT=fastap -SUFFIX=olip2n.rng -GAPOP=4.500
 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality
 -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pts -NORM=ext -MINLEN=0
 -MAXLEN=2000000000 -USER=US09126945_@CGN1_1_398 -NCPU=6 -ICPU=3
 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-126-945B-2_COPY_84_92
Query length: 9
Database: N_Geneseq_0601:*
Database sequences: 730101
Database length: 313950809
Search time (sec): 680.150000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000


score_list:
Sequence                   Strd Orig    ZScore    EScore Len  ! Documentation ..
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF21828 +   [illegible]  [illegible]  [illegible]  1087 !
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:[illegible] +   [illegible]  [illegible]  [illegible]  [illegible] !
/SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:AA[illegible]261 +   9.00  166.38  0.2690  1894 !
/SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:AA[illegible]266 +   9.00  162.91  0.4252  3217 !
/SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT:AA[illegible]508 +   6.00  121.22  59.93  234 !
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AA[illegible] +   6.00  121.22  [illegible]  [illegible] !
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA7[illegible] +   [illegible]  121.22  [illegible]  [illegible] !
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA36810 +   6.00  123.48  65.93  263 !
/SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:AAX[illegible] +   [illegible]  [illegible]  [illegible]  263 !
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AA[illegible]181 +   [illegible]  [illegible]  [illegible]  [illegible] !
/SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT:AA[illegible] +   6.00  119.91  104.28  461 !
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AA[illegible] +   6.00  119.91  104.28  461 !
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA7[illegible] +   6.00  119.91  104.28  461 !
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX[illegible] +   [illegible]  118.97  117.58  534 !
/SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV[illegible] +   [illegible]  [illegible]  [illegible]  [illegible] !
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA[illegible] +   [illegible]  [illegible]  [illegible]  [illegible] !
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AA[illegible]427 +   6.00  117.66  [illegible]  [illegible] !
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AA[illegible]58 +   6.00  [illegible]  [illegible]  [illegible] !
/SIDS1/gcgdata/geneseq/geneseqn/NA1989.DAT:AAN[illegible] +   6.00  115.47  184.[illegible]  [illegible] !
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AA[illegible] +   [illegible]  115.[illegible]  188.[illegible]  [illegible] !
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX[illegible]198 +   6.00  114.65  204.56  1052 !
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX13745 +   6.00  113.73  230.41  1217 !
/SIDS1/gcgdata/geneseq/geneseqn/NA1995.DAT:AA[illegible]372 -   6.00  113.47  238.11  1267 !
/SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV27[illegible] +   [illegible]  113.47  238.11  1267 !
/SIDS1/gcgdata/geneseq/geneseqn/NA1992.DAT:AA[illegible] +   [illegible]  [illegible]  [illegible]  [illegible] !
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AA[illegible]384 +   [illegible]  [illegible]  247.71  [illegible] !
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AA[illegible]6152 +   6.00  112.75  261.20  1419 !
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAY1[illegible] +   [illegible]  111.1[illegible]  [illegible].42  [illegible] !
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AA[illegible] +   [illegible]  [illegible]  [illegible]  [illegible] !
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA90[illegible] +   6.00  110.78  336.07  [illegible] !
/SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV11854 -   6.00  106.96  548.94  3523 !
/SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV[illegible] +   [illegible]  [illegible]  [illegible]  [illegible] !
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA[illegible] +   [illegible]  [illegible]  567.7[illegible]  [illegible] !
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA[illegible] +   [illegible]  106.6[illegible]  567.7[illegible]  3671 !
/SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF54[illegible] +   6.00  106.69  567.70  3671 !
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA36745 +   [illegible]  102.66  [illegible]  [illegible] !
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAX[illegible] +   [illegible]  [illegible]  [illegible]  [illegible] !
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA[illegible] +   [illegible]  102.[illegible]  975.57  6240 !
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAX82946 +   6.00  102.[illegible]  975.57  6240 !
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAX36757 +   6.00  102.[illegible]  951.63  6910
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA36763 +   6.00  102.[illegible]  951.63  6910
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAX82848 +   6.00  102.[illegible]  951.63  6910
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAX82854 +   6.00  102.67  951.63  6910
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA36744 +   6.00  102.67  951.74  6911
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA36756 +   6.00  102.67  951.74  6911

seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF21828

seq_documentation_block:
ID   AAF21828 standard; DNA; 1087 BP.
XX
AC   AAF21828;
XX
DT   27-MAR-2001  (first entry)
XX
DE   Human breast and ovarian cancer associated antigen gene SEQ ID 215.
XX
KW   Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
KW   nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW   antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
KW   antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
KW   Addison's disease; allergy; autoimmune haemolytic anaemia;
KW   autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW   multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW   cardiovascular disorder; wound healing; neurological disease; ss.
XX
OS   Homo sapiens
XX
PN   WO200055173-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05881.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-611515/58.
DR   P-PSDB; AAB58925.
XX
PT   New human breast and ovarian cancer associated gene sequences and the
PT   polypeptides encoded by these genes, useful in the prevention,
PT   treatment and diagnosis of cancer, immune disorders, cardiovascular
PT   disorders and neurological diseases -
XX
PS   Claim 13; Page 642-643; 1299pp; English.
XX
CC   Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
CC   proteins AAB58711 - AAB59128. The DNA and protein sequences are
CC   associated with breast and ovarian cancer. Included in the invention are
CC   sequences AAF22032 - AAF22040 and AAB59129 which are used in the
CC   isolation and characterisation of the DNA and protein sequences of the
CC   invention. The breast and ovarian cancer associated DNA, protein, agonist
CC   [illegible] sequences exhibit cytostatic, immunosuppressive;
CC   nootropic, neuroprotective, antiviral, antiallergic, hepatotropic;
CC   antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
CC   antibacterial; antifungal; antiparasitic and cardiant activity. The
CC   polynucleotide and protein sequences are used in the diagnosis of cancer,
CC   particularly breast and ovarian cancer. The nucleic acid sequences,
CC   proteins, agonists and antagonists may also be used in the diagnosis,
CC   prevention and treatment of immune disorders e.g. Addison's disease,
CC   allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC   diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC   arthritis and ulcerative colitis, cardiovascular disorders such as
CC   myocardial ischaemias; wound healing; neurological diseases such as
CC   cerebral [illegible] and epilepsy; and infectious diseases.
XX
SQ   Sequence 1087 BP; 215 A; 417 C; 330 G; 182 T; 4 other;
```

```
Alignment Scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment Block:
 US-09-126-945B-2_COPY_84_92 x AAF21828  ..

 Align Seq 1/1  to: AAF21828  from: 1  to: 1097

        1 GluGluProProGluGluProGluGln 9
          ||||||||||||||||||||||||||||
      126 GAGGAGCCACCTGAGGAGCCTGAGCAG 152

Seq Name:  %title[period]type seq species pg NA...ctLA..AA..50691

Seq Documentation Block:
ID   AAZ50691 standard; cDNA; 1894 BP.
XX
AC   AAZ50691;
XX
DT   23-MAY-2000  (first entry)
XX
DE   Human prostate derived Ets factor cDNA.
XX
KW   Prostate derived Ets factor; PDEF; human; chromosome 6p21.3; cancer;
KW   loss of heterozygosity; chromosomal translocation; linkage analysis;
KW   cytostatic; cardiant; immunosuppressive; cerebroprotective; fungicide;
KW   antibacterial; vulnerary; neuroprotective; antiparkinsonian; nootropic;
KW   anabolic; antiinflammatory; anorectic; hybridisation probe; forensic;
KW   tumour marker, diagnosis, treatment, prostate cancer, blood coagulation,
KW   autoimmune disorder, haematopoietic, immunosuppressive system, stroke;
KW   neoplasm; microbial infection; tissue regeneration; heart attack;
KW   scarring; food additive; preservative; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             416..1423
FT                   /*tag= a
FT                   /product= "Human prostate derived Ets factor"
FT                   /note= "Expressed in prostate epithelium"
XX
PN   WO200006589-A1.
XX
PD   10-FEB-2000.
XX
PF   02-AUG-1999;   99WO-US17470.
XX
PR   31-JUL-1998;   98US-0126945.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI   Libermann TA, Oettgen JP, Kunsch CA, Endress GA, Rosen CA;
XX
DR   WPI; 2000-195255/17.
DR   P-PSDB; AAY45003.
XX
PT   Novel prostate derived polypeptide, polynucleotide useful for
PT   diagnosis, prevention and treatment of prostate cancer, autoimmune
PT   disorders, microbial infections and also as food additive or
PT   preservative
XX
PS   Claim 4; Fig 1; 142pp; English.
XX
CC   The present cDNA sequence encodes a human prostate derived Ets factor
CC   (PDEF), isolated from human prostate carcinoma cDNA library
CC   (ATCC No.203072). PDEF gene expression is highest in tissues involved in
CC   androgen sensitivity, with enriched expression in prostate epithelium
CC   cells. PDEF gene is mapped to the human chromosome 6p21.3
CC   region that is associated with loss of heterozygosity and chromosomal
CC   translocations in various human cancers. PDEF has cytostatic, cardiant,
CC   immunosuppressive, cerebroprotective, fungicide, antibacterial,
CC   vulnerary, neuroprotective, antiparkinsonian, nootropic, anabolic,
CC   antiinflammatory and anorectic activity. PDEF polynucleotides are useful
CC   in linkage analysis as markers, as hybridisation probes for differential
CC   identification of the tissues or cell types and as polymorphic markers
CC   for forensic purposes. PDEF is useful as prostate specific tumour marker
CC   for the diagnosis and treatment of prostate cancer. PDEF sequences are
CC   useful for treating autoimmune disorders, haematopoietic, blood
CC   coagulation, immune and nervous system disorders, hyperproliferative
CC   disorders like, neoplasms and microbial infections, heart attacks,
CC   stroke, scarring and for tissue regeneration. They are also useful as
CC   food additives or preservatives.
XX
SQ   Sequence 1894 BP; 468 A; 654 C; 571 G; 302 T; 0 other;


Alignment Scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment Block:
 US-09-126-945B-2_COPY_84_92 x AAZ50691  ..

 Align Seq 1/1  to: AAZ50691  from: 1  to: 1894

        1 GluGluProProGluGluProGluGln 9
          ||||||||||||||||||||||||||||
      665 GAGGAGCCACCTGAGGAGCCTGAGCAG 691

Seq Name:  %title[period]type seq species pg NA2001.DAT:AAC83261

Seq Documentation Block:
ID   AAC83261 standard; DNA; 1894 BP.
XX
AC   AAC83261;
XX
DT   16-MAR-2001  (first entry)
XX
DE   Gland-specific Ets transcription factor (GSEF) nucleotide sequence.
XX
KW   Transcription factor; gland specific Ets transcription factor; GSEF;
KW   metastatic potential; cancer; tumour; metastasis; breast; prostate;
KW   leukaemia; lymphoma; sarcoma; melanoma; chromosome 6p21.1 6p21.3; ds
XX
OS   Homo sapiens.
XX
PN   WO200020592-A1.
XX
PD   23-NOV-2000.
XX
PF   12-MAY-2000;  2000WO-US13173.
XX
PR   14-MAY-1999;   99US-0134112.
XX
PA   (CHIR-) CHIRON CORP.
XX
PI   Kaufmann J, Xin H, Harrowe G;
XX
DR   WPI; 2001-041979/05.
DR   P-PSDB; AAB49628.
XX
PT   Detecting metastatic and potential metastatic cancerous cells, useful
PT   for diagnosing, prognosing, grading and staging of cancers by detecting
PT   gland specific Ets transcription factor gene product in a biological
PT   sample from a cell
XX
PS   Disclosure; Page 81-83; 95pp; English.
XX
CC   This invention relates to a method for the detection and determination of
CC   the metastatic potential of a cell. The method comprises detecting a
CC   gland specific Ets transcription factor (GSEF) gene product in a test
```

```
XX   sample. Detection of a GSEF gene product in the test sample in amount
CC   lower than that in a normal cell, is indicative of a cell with high
CC   metastatic potential. The method is useful for determining the metastatic
CC   potential of a cell, for the diagnosis and prognosis of cancer as well as
CC   grading and staging of cancers by detecting GSEF expression in a
CC   biological test sample. The method may also be used to monitor patients
CC   having a predisposition to develop a particular cancer. GSEF polypeptides
CC   are useful for producing antibodies, in cancer diagnosis, prognosis,
CC   grading, staging and management of breast and prostate tumours, and in
CC   detecting polymorphisms in the sequence. GSEF genes and proteins are also
CC   useful in gene therapy. GSEF gene product expression levels can be used
CC   in conjunction with any tissue in which an alteration in GSEF gene
CC   product expression levels is associated with development of a
CC   cancer-associated phenotype. Cancers, which can be monitored include
CC   cancers of the prostate, cervix, lung and colon, melanomas, colorectal
CC   adenocarcinomas, Wilm's tumour, retinoblastoma, sarcomas, myosarcomas,
CC   lung carcinomas, leukaemia, and lymphomas. The GSEF gene is located on
CC   human chromosome 6, specifically at 6p21.1-6p21.3. The present sequence
CC   represents the DNA encoding GSEF.
XX
SQ   Sequence 1894 BP; 368 A; 653 C; 571 G; 302 T; 0 other;


alignment_scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_84_92 x AAC83261     .

 Align seq 1/1  to: AAC83261  from: 1  to: 1894

        1 GluGluProProGluGluProGluGln 9
          |||||||||||||||||||||||||||
      665 GAGGAGCCACCTGAGGAGCCTGAGCAG 691

Seq_name: /cgsi/ie_data/geneseq_genes/geneseq_NA1996.DAT:AAC83266

Seq_documentation_block:
ID   AAC83266 standard; DNA; 3317 BP.
XX
AC   AAC83266;
XX
DT   16-MAR-2001  (first entry)
XX
DE   Gland-specific Ets transcription factor (GSEF) cDNA sequence.
XX
KW   Transcription factor; gland-specific Ets transcription factor; GSEF;
KW   metastatic potential; cancer; tumour; metastasis; breast; prostate;
KW   leukaemia; lymphoma; sarcoma; melanoma; chromosome 6p21.1-6p21.3; ss.
XX
OS   Homo sapiens.
XX
PN   WO200070092-A1.
XX
PD   23-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US13173.
XX
PR   14-MAY-1999;   99US-0134112.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Kaufmann J,  Xin H,  Harrower G;
XX
DR   WPI; 2001-041019/05.
DR   P-PSDB; AAB49633.
XX
PT   Detecting metastatic and potential metastatic cancerous cells, useful
PI   for diagnosing, prognosing, grading and staging of cancers by detecting
PI   gland-specific Ets transcription factor gene product in a biological
PI   sample from a cell -
XX
PS   Disclosure; Fig 2; 95pp; English.
XX
CC   This invention relates to a method for the detection and determination of
CC   the metastatic potential of a cell. The method comprises detecting a
CC   gland-specific Ets transcription factor (GSEF) gene product in a test
CC   sample. Detection of a GSEF gene product in the test sample in amount
CC   lower than that in a normal cell, is indicative of a cell with high
CC   metastatic potential. The method is useful for determining the metastatic
CC   potential of a cell, for the diagnosis and prognosis of cancer as well as
CC   grading and staging of cancers by detecting GSEF expression in a
CC   biological test sample. The method may also be used to monitor patients
CC   having a predisposition to develop a particular cancer. GSEF polypeptides
CC   are useful for producing antibodies, in cancer diagnosis, prognosis,
CC   grading, staging and management of breast and prostate tumours, and in
CC   detecting polymorphisms in the sequence. GSEF genes and proteins are also
CC   useful in gene therapy. GSEF gene product expression levels can be used
CC   in conjunction with any tissue in which an alteration in GSEF gene
CC   product expression levels is associated with development of a
CC   cancer-associated phenotype. Cancers, which can be monitored include
CC   cancers of the prostate, cervix, lung and colon, melanomas, colorectal
CC   adenocarcinomas, Wilm's tumour, retinoblastoma, sarcomas, myosarcomas,
CC   lung carcinomas, leukaemia, and lymphomas. The GSEF gene is located on
CC   human chromosome 6, specifically at 6p21.1-6p21.3. The present sequence
CC   represents the cDNA encoding GSEF.
XX
SQ   Sequence 3317 BP; 716 A; 1026 C; 970 G; 611 T; 0 other;


alignment_scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_84_92 x AAC83266     .

 Align seq 1/1  to: AAC83266  from: 1  to: 3317

        1 GluGluProProGluGluProGluGln 9
          |||||||||||||||||||||||||||
     2064 GAGGAGCCACCTGAGGAGCCTGAGCAG 2090

Seq_name: /cgsi/ie_data/geneseq_genes/geneseq_NA1996.DAT:AAI18508

Seq_documentation_block:
ID   AAI18508 standard; DNA; 234 BP.
XX
AC   AAI18508;
XX
DT   17-NOV-1996  (first entry)
XX
DE   NAP probe derived from AcaNAP5.
XX
KW   AcaNAP; BpaNAP; NamNAP; AceNAP; AdhNAP; anticoagulant;
KW   nematode extracted anticoagulant protein; serine protease;
KW   nematode; thrombosis; parasitic worm; ss.
XX
OS   Synthetic.
XX
PN   WO9612021-A2.
XX
PD   25-APR-1996.
XX
PF   17-OCT-1995;   95WO-US13231.
XX
PR   05-JUN-1995;   95US-0460399.
PR   18-OCT-1994;   94US-0326110.
PR   05-JUN-1995;   95US-0461965.
PR   05-JUN-1995;   95US-0465380.
PR   05-JUN-1995;   95US-0486397.
XX
PA   (CORV-) CORVAS INT INC.
```

XX
PI Bergum PW, Gansemans YGJ, Jespers LS, Laroche YR;
PI Lauwereys MJ, Messens JHL, Moyle M, Stanssens PEH;
PI Vlasuk GP;
XX
DR WPI; 1996-222097/22.
XX
PT Proteins with anticoagulant and/or serine protease inhibitory
PT activity isolated from nematodes and useful to inhibit blood
PT coagulation
XX
PS Disclosure; Page 57; 243pp; English.
XX
CC Proteins with anticoagulant and/or serine protease inhibitory
CC activity, isolated from nematodes, are useful to inhibit blood
CC coagulation. The proteins can be added to blood collection tubes
CC defining the collection of mammalian plasma. They are also useful
CC to prevent or inhibit thrombosis, and may be given alone or in
CC combination with other therapeutic or in vivo diagnostic agents.
CC The proteins can serve as immunogens to raise antibodies for use in
CC the diagnosis and identification of NAP concn. levels in biological
CC fluids, e.g. to detect mammalian infection with a parasitic worm.
CC They can also be used as immunogens in prophylactic and therapeutic
CC vaccines against parasitic worm infection. The proteins may
CC double the clotting time of human plasma in prothrombin time assays
CC when present at 10-50 nMol, and double the clotting time of human
CC plasma in activated partial thrombin time assays when present
CC at 10-100 nMol.
CC The anticoagulant proteins are pref. derived from
CC Ancylostoma caninum, A. ceylanicum, A. duodenale, Necator
CC americanus or Heligmosmoides polygyrus.
CC The proteins pref. have 2 NAP domains and specifically inhibit
CC the catalytic activity of the factor VIIa/TF complex in the
CC presence of factor Xa or a catalytically inactive factor Xa deriv.,
CC do not specifically inhibit the activation of factor VIIa in the
CC absence of TF and do not specifically inhibit prothrombinase.
XX
SQ Sequence 234 BP; 66 A; 49 C; 70 G; 49 T; 0 other;

alignment scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2_COPY_84_92 x AAT18508 ..

Align seq 1/1 to: AAT18508 from: 1 to: 234

1 GluGluProProGluGlu 6
||||||||||||||||||
79 GAGGAACCCCCTGAGGAG 96

Seq name: /tmp/biosci/data/geneseq/geneseqp-embl.seq:AAT18508 AA:AA;04;*

seq documentation block:
ID AAZ10437 standard; cDNA; 234 BP.
XX
AC AAZ10437;
XX
DT 15-NOV-1999 (first entry)
XX
DE Probe derived from cDNA encoding a NAP.
XX
KW Nematode extracted anticoagulant protein; NAP; anticoagulant;
KW serine protease inhibitor; NAP domain; factor VIIa/TF; probe; ss.
XX
OS Ancylostoma caninum.
XX
PN US5955294-A.
XX
PD 21-SEP-1999.

XX
PF 19-APR-1996; 96US-0634641.
XX
PR 19-APR-1996; 96US-0634641.
PR 18-OCT-1994; 94US-0326110.
PR 05-JUN-1995; 95US-0461965.
PR 05-JUN-1995; 95US-0465380.
PR 05-JUN-1995; 95US-0486397.
PR 05-JUN-1995; 95US-0486399.
PR 17-OCT-1995; 95WO-US13231.
XX
PA (COVX) COVAS INT INC.
XX
PI Bergum PW, Gansemans YGJ, Jespers LS, Laroche YR;
PI Lauwereys MJ, Messens JHL, Moyle M, Stanssens PEH;
PI Vlasuk GP;
XX
DR WPI; 1999-539969/45.
XX
PT Screening an isolated protein for Nematode extracted Anticoagulant
PT Protein domains
XX
PS Disclosure; Columns 85-86; 197pp; English.
XX
CC The present sequence is a probe derived from cDNA encoding a nematode
CC extracted anticoagulant protein (NAP). The protein has activity as an
CC anticoagulant and/or serine protease inhibitor. The protein contains
CC at least one NAP domain which has selective inhibitory activity for
CC factor VIIa/TF. The specification describes a method for screening an
CC isolated protein at least one domain for factor VIIa/TF selective
CC inhibitory activity. The method comprises determining the time to
CC clotting effected by a concentration of the isolated protein in an
CC ex- prothrombin time (PT) assay and in vitro activated partial
CC thromboplastin time (aPTT) assay; calculating prolongation of clotting
CC effected by the isolated protein in each of the PT and aPTT assay, with
CC respect to a baseline clotting value for each assay, where prolongation
CC of clotting is calculated as fold elevation of clotting time relative
CC to a baseline clotting value, where a doubling of clotting time is
CC deemed a two-fold elevation; and calculating a PI to aPTT prolongation
CC ratio, where a ratio at least one is indicative of factor VIIa/TF
CC inhibitory activity. The method is useful for determining if a protein
CC has factor VIIa/TF inhibitory activity.
XX
SQ Sequence 234 BP; 66 A; 49 C; 70 G; 49 T; 0 other;

alignment scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2_COPY_84_92 x AAZ10437 ..

Align seq 1/1 to: AAZ10437 from: 1 to: 234

1 GluGluProProGluGlu 6
||||||||||||||||||
79 GAGGAACCCCCTGAGGAG 96

Seq name: /tmp/biosci/data/geneseq/GENESEQP-NEW.DAT:AAZ10437

seq documentation block:
ID AAZ10437 standard; cDNA; 234 BP.
XX
AC AAZ10437;
XX
DT 19-DEC-2000 (first entry)
XX
DE Nematode extracted anticoagulant proteins hybridisation probe #1.
XX
KW Nematode extracted anticoagulant protein; blood clotting; thrombosis;
KW canine hookworm; vaccine; probe; ss.

```
XX
OS   Unidentified.
XX
PN   US6087487-A.
XX
PD   11-JUL-2000.
XX
PF   12-FEB-1999;   99US-0249451.
XX
PR   17-OCT-1995;   95WO-US13231.
PR   17-APR-1997;   97US-0809455.
PR   18-OCT-1994;   94US-0326110.
PR   05-JUN-1995;   95US-0461965.
PR   05-JUN-1995;   95US-0465380.
PR   05-JUN-1995;   95US-0486397.
PR   05-JUN-1995;   95US-0486399.
XX
PA   (CORV-) CORVAS INT INC.
XX
PI   Lauwereys MJ,  Stanssens PEH,  Jespers LS,  Lansemans YGI,  Moyle M,
PI   Bergum PW,  Messens JHL,  Lakeche YE,  Vlasuk GP;
XX
DR   WPI; 2000-541359/48.
XX
PT   New cDNA molecule encoding a protein having factor Xa inhibitory
PT   activity for preventing and treating blood clotting disorders,
PT   comprises nematode-extracted anticoagulant protein domains -
XX
PS   Disclosure; Column 5; 197pp; English.
XX
CC   The present sequence is a probe used to isolate a number of cDNA
CC   sequences encoding nematode-extracted anticoagulant proteins (NAPs). NAPs
CC   have been shown to be effective at preventing blood clotting without
CC   causing excessive bleeding. The proteins can be used in blood collection
CC   tubes to aid the isolation of plasma from the blood, to prevent
CC   thrombosis which may be linked to the rupture of an atherosclerotic
CC   plaque, acute myocardial infarction, angina, thrombolytic therapy,
CC   percutaneous transluminal coronary angioplasty, disseminated
CC   intravascular coagulopathy, infection, cancer and septic shock, and to
CC   produce antibodies. In the latter instance, the antibodies can be raised
CC   in order to detect infection by nematodes (the coding sequences can also
CC   be used for this) or as diagnostic tests. The proteins can also be used
CC   as vaccines against nematode parasites.
XX
SQ   Sequence 244 BP; 66 A; 49 C; 70 G; 49 T; 0 other;


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_84_92 x AAA73360    ..

 Align seq 1/1   to: AAA73360   from: 1   to: 244

       1 GluGluProProGluGlu 6
         ||||||||||||||||||
      79 GAGGAACCCCCTGAGGAG 96

seq_name:  "sISS:Translated genes proteins;NA2700:CAT:AAA36810

seq_documentation_block:
ID   AAA36810 standard; DNA; 264 BP.
XX
AC   AAA36810;
XX
DT   03-AUG-2000  (first entry)
XX
DE   Human dysterlin related nucleotide sequence SEQ ID NO:68.
XX
KW   Human; dysterlin; mutant; identification; chromosome 2p12-14;
KW   detection; muscular dystrophy; diagnosis; hereditary muscular dystrophy;
KW   miyoshi myopathy; limb girdle muscular dystrophy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200011016-A1.
XX
PD   02-MAR-2000.
XX
PF   25-AUG-1999;   99WO-US19394.
XX
PR   25-AUG-1998;   98US-0097930.
XX
PA   (GEBO-) GEN HOSPITAL CORP.
PA   (UYPI-) UNIV PITTSBURGH.
XX
PI   Brown RH,  Liu J,  Hoffman E,  Chou F;
XX
DR   WPI; 2000-246531/21.
XX
PT   Dysferlin polynucleotide, its mutant form useful for diagnosis and
PT   treatment of hereditary muscular dystrophies e.g. miyoshi myopathy and
PT   limb girdle muscular dystrophy -
XX
PS   Disclosure; Page 105; 136pp; English.
XX
CC   The present invention describes an isolated dysferlin DNA of >= 25
CC   nucleotides in length, comprising a nucleotide sequence specifically
CC   selected from nucleotides 911-913, 929-948, 1019-1038, 1392-1411,
CC   1424-1443, 1484-1503, 1499-1518, 1543-1565, 1715-1734, 1714-1733,
CC   2241-2260, 2864-2883, 2978-2997, 3057-3076, 3198-4217, 4252-4271,
CC   4356-4375, 4665-4684, 5015-5034, 5610-5629, 5726-5745, 6035-6054,
CC   6179-6198, 6243-6263 and 6529-6548 of the human dysferlin nucleotide
CC   sequence given in AAA36744. Dysferlin nucleotide sequences containing
CC   specific mutations can be used for diagnosing a patient, a foetus or
CC   a pre-embryo at risk of developing a dysferlin associated disorder by
CC   detecting mutations in the dysferlin gene in biological samples from
CC   patient's. Alternatively, the biological sample containing genomic DNA
CC   can be incubated with a restriction enzyme, preferably BanII, Bspl286I,
CC   RsaI, HhaI, BaeIII, Bspl286, NlaIV, NlaIII, BcuI, AvaII, BstEII, PleI,
CC   HaeI, AluI, ApoI, Tsp509I, SalI, HincII, TaqI, HinfI, TliI, StuNI or
CC   FokI and the presence or absence of a restriction enzyme site in the
CC   sample is detected as an indication of the presence or absence of a
CC   particular mutation in the sample. Dysferlin polynucleotides are useful
CC   for treating hereditary muscular dystrophies such as miyoshi myopathy
CC   (MM) and limb girdle muscular dystrophy-2B (LGMD-2B). MM and LGMD-2B
CC   map to the human chromosome 2p12-14 region between the genetic markers
CC   D2S292 and D2S286. The present sequence represents a nucleotide sequence
CC   given in the sequence listing of the present specification, but which
CC   does not appear to be mentioned further.
XX
SQ   Sequence 264 BP; 50 A; 85 C; 67 G; 61 T; 0 other;


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_84_92 x AAA36810    ..

 Align seq 1/1   to: AAA36810   from: 1   to: 264

       2 GluProProGluGluPro 7
         |||||||||||||||||
     153 GAACCTCCGGAAGAACCC 170

seq_name:  "sISS:Translated genes proteins;NA2700:CAT:AAA82902

seq_documentation_block:
ID   AAA82902 standard; DNA; 264 BP.
XX
```

```
AC   AAX82901;
XX
DT   30-JUN-2000  (first entry)
XX
DE   Human dysferlin DNA #49.
XX
KW   Dysferlin; anti dystrophic; gene therapy; muscular dystrophy; human;
KW   skeletal muscle cell; hereditary; Miyoshi myopathy; diagnosis;
KW   limb girdle muscular dystrophy 2B; brain specific; ss.
XX
OS   Homo sapiens.
XX
PN   WO200011157-A1.
XX
PD   02-MAR-2000.
XX
PF   25-AUG-1999;   99WO-US019395
XX
PR   25-AUG-1998;   98US-0097927.
XX
PA   (GEBO ) GEN HOSPITAL CORP.
XX
PI   Brown RH, Liu J, Aoki M, Ho MF, Matsuda-Asada C;
XX
DR   WPI; 2000-237646/20.
XX
PT   Novel dysferlin genes and related proteins useful for diagnosis, risk
PT   identification and treatment of hereditary muscular dystrophies and
PT   other dysferlin related disorders
XX
PS   Claim 11; Page 115; 146pp; English.
XX
CC   This invention describes a novel human dysferlin nucleic acid (I) and
CC   its encoding protein (II), which has anti-dystrophic activity and can be
CC   used for gene therapy. Introduction of (I), a vector comprising (I) or
CC   dysferlin into a cell of a mammal can be used to decrease the symptoms of
CC   muscular dystrophy. The dysferlin gene is normally expressed in skeletal
CC   muscle cells and is selectively mutated in several families with the
CC   hereditary muscular dystrophies, e.g. Miyoshi myopathy and limb girdle
CC   muscular dystrophy-2B. The primers and oligonucleotides derived from (I)
CC   can be used in diagnosis of or risk identification for dysferlin related
CC   disorders in patients, tetus, or pre-embryos. Expression of
CC   brain specific dysferlin may be important as a marker for normal neural
CC   development. Dysferlin DNA or subgenomic fragments can be used in
CC   therapy of the hereditary muscular dystrophies. This sequence represents
CC   a fragment of the human dysferlin gene described in the method of the
CC   invention.
XX
SQ   Sequence 264 BP; 50 A; 85 C; 67 G; 61 T; 0 other;


alignment scores:
                Quality:     6.00              Length:       6
                  Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_84_92 x AAX82901  ..

 Align seq 1/1  to: AAX82901  from: 1  to: 264

        2 GluProProGluGluPro 7
          ||||||||||||||||||
      154 GAACCTCCCGAAGAACCC 170

Seq name: /cgi-bin/dirdb/genseq-fetch.pl?pn|NA2000|DAI:AAX82901

Seq documentation block:
ID   AAX82901 standard; DNA; 408 BP.
XX
AC   AAX82901;
XX
DT   25-JAN-2001  (first entry)
XX
DE   Eucalyptus grandis transcription factor DNA sequence #455.
XX
KW   Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW   poplar; sweetgum; teak; mahogany; bZIP; G box binding factor;
KW   basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW   homeodomain zipper; LIM domain; AP2; ERE8P; zinc finger domain;
KW   type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
OS   Eucalyptus grandis.
XX
PN   WO200057724-A2.
XX
PD   14-SEP-2000.
XX
PF   09-MAR-2000; 2000WO-US006112.
XX
PR   11-MAR-1999;   99US-0266513.
PR   18-AUG-1999;   99US-0149485.
XX
PA   (GENE ) GENESIS RES & DEV CORP LTD.
PA   (FLEI ) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Wood M, McGrath A, Shenk MA, Glenn M;
XX
DR   WPI; 2000-579869/54.
XX
PT   New isolated polynucleotide encoding a plant transcription factor for
PT   production a plant e.g. a woody plant, preferably eucalyptus or pine,
PT   having modified gene expression or modified activity of a polypeptide
PT   .
XX
PS   Claim 1; Page 446; 747pp; English.
XX
CC   The present invention relates to novel plant transcription factors from
CC   Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC   sequence for one such transcription factor. The transcription factor may
CC   be used to produce a plant having modified gene expression such as a
CC   woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC   mahogany species or to modify the activity of a polypeptide in a plant.
CC   The transcription factors of the present invention are members from the
CC   following families of regulatory proteins: bZIP, bZIP family, basic
CC   binding factors, basic helix-loop-helix zipper,
CC   homeotic, homeodomain, homeobox, MADS, homeodomain zipper, LIM domain, AP2
CC   and ERE8P, zinc finger domains of type 2 Cys2His2, CCAAT box elements
CC   and MYB.
XX
SQ   Sequence 408 BP; 83 A; 136 C; 129 G; 59 T; 1 other;


alignment scores:
                Quality:     6.00              Length:       6
                  Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_84_92 x AAC56484  ..

 Align seq 1/1  to: AAC56484  from: 1  to: 408

        4 ProGluGluProGluGln 9
          ||||||||||||||||||
      283 CCAGAAGAGCCAGAGCAA 300

Seq name: /cgi-bin/dirdb/genseq-fetch.pl?pn|NA1996.DAI:AAI12945

Seq documentation block:
ID   AAI12945 standard; cDNA; 461 BP.
XX
AC   AAI12945;
XX
DT   17-NOV-1999  (first entry)
XX
```

```
DE   AcaNAP5.
XX
KW   AcaNAP; BpoNAP; NamNAP; AceNAP; AduNAP; anticoagulant;
KW   nematode extracted anticoagulant protein; serine protease;
KW   nematode; thrombosis; parasitic worm; ss.
XX
OS   Ancylostoma caninum.
XX
FH   Key             Location/Qualifiers
FT   CDS             22..324
FT                   /*tag   a
FT                   /label  AcaNAP5
FT   misc_RNA        79..321
FT                   /*tag   b
FT                   /label  pro-AcaNAP5
FT   mat_peptide     91..321
FT                   /*tag   c
XX
PN   WO9612021-A2.
XX
PD   25-APR-1996.
XX
PF   17-OCT-1995;    95WO-US13231.
XX
PR   05-JUN-1995;    95US-0486399.
PR   18-OCT-1994;    94US-0326110.
PR   05-JUN-1995;    95US-0461965.
PR   05-JUN-1995;    95US-0465380.
PR   05-JUN-1995;    95US-0486397.
XX
PA   (CORV-) CORVAS INT INC.
XX
PI   Bergum PW,  Gansemans YGJ,  Jospers LS,  Laroche YR;
PI   Lauwereys MJ,  Messensjhl,  Moyle M,  Stanssens PEH;
PI   Vlasuk GP;
XX
DR   WPI; 1996-222007/22.
DR   P-PSDB; AAR91699.
XX
PT   Proteins with anticoagulant and/or serine protease inhibitory
PT   activity - isolated from nematodes and useful to inhibit blood
PT   coagulation
XX
PS   Claim 46; Fig 1; 243pp; English.
XX
CC   Proteins with anticoagulant and/or serine protease inhibitory
CC   activity, isolated from nematodes, are useful to inhibit blood
CC   coagulation.  The proteins can be added to blood collection tubes
CC   defining the collection of mammalian plasma.  They are also useful
CC   to prevent or inhibit thrombosis, and may be given alone or in
CC   combination with other therapeutic or in vivo diagnostic agents.
CC   The proteins can serve as immunogens to raise antibodies for use in
CC   the diagnosis and identification of NAP concn. levels in biological
CC   fluids, e.g. to detect mammalian infection with a parasitic worm.
CC   They can also be used as immunogens in prophylactic and therapeutic
CC   vaccines against parasitic worm infection.  The proteins may
CC   double the clotting time of human plasma in prothrombin time assays
CC   when present at 10-50 nMol, and double the clotting time of human
CC   plasma in activated partial thrombin time assays when present
CC   at 10-100 nMol.
CC   The anticoagulant proteins are pref. derived from
CC   Ancylostoma caninum, A. ceylanicum, A. duodenale, Necator
CC   americanus or Heligomosmoides polygyrus.
CC   The proteins pref. have 2 NAP domains and specifically inhibit
CC   the catalytic activity of the factor VIIa/TF complex in the
CC   presence of factor Xa or a catalytically inactive factor Xa deriv.,
CC   do not specifically inhibit the activation of factor VIIa in the
CC   absence of TF and do not specifically inhibit prothrombinase.
XX
SQ   Sequence 461 BP; 134 A; 99 C; 115 G; 113 T; 0 other;


alignment_scores:
                Quality:     6.00              Length:       6
                  Ratio:    1.000               Gaps:        0
  Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_84_92 x AAI12945    ..

 Align Seq 1/1   to: AAI12945   from: 1   to: 461

        1 GlnGluProProGlnGln 6
          ||||||||||||||||||
      169 GAGGAACCCCCTCAGGAG 186

seq_guide  /TF  Ligand/Factor Xa/prothrombin time (PT) assay/activated partial thromboplastin time (aPTT) assay/AAI12945

Seq_documentation_block:
ID   AAZ10439 standard; cDNA; 461 BP.
XX
AC   AAZ10439;
XX
DT   15-NOV-1999  (first entry)
XX
DE   cDNA encoding nematode extracted anticoagulant protein AcaNAP5.
XX
KW   Nematode extracted anticoagulant protein; NAP; anticoagulant;
KW   serine protease inhibitor; NAP domain; factor VIIa/TF; ss.
XX
OS   Ancylostoma caninum.
XX
PN   US5955294-A.
XX
PD   21-SEP-1999.
XX
PF   19-APR-1996;    96US-0634641.
XX
PR   19-APR-1996;    96US-0634641.
PR   18-OCT-1994;    94US-0326110.
PR   05-JUN-1995;    95US-0461965.
PR   05-JUN-1995;    95US-0465380.
PR   05-JUN-1995;    95US-0486397.
PR   05-JUN-1995;    95US-0486399.
PR   17-OCT-1995;    95WO-US13231.
XX
PA   (CORV-) CORVAS INT INC.
XX
PI   Bergum PW;  Gansemans YGJ;  Jospers LS;  Laroche YR;
PI   Lauwereys MJ;  Messens JHL;  Moyle M;  Stanssens PEH;
PI   Vlasuk GP;
XX
DR   WPI; 1999-539569/45.
DR   P-PSDB; AAY30382.
XX
PT   Screening an isolated protein for Nematode extracted Anticoagulant
PT   Protein domains
XX
PS   Disclosure; Fig 1; 197pp; English.
XX
CC   The present sequence encodes a nematode extracted anticoagulant protein
CC   (NAP). The protein has activity as an anticoagulant and/or serine
CC   protease inhibitor. The protein contains at least one NAP domain which
CC   has selective inhibitory activity for factor VIIa/TF. The specification
CC   describes a method for screening an isolated protein at least one domain
CC   for factor VIIa/TF selective inhibitory activity. The method comprises
CC   determining the time to clotting effected by a concentration of the
CC   isolated protein in a 1 x 5% prothrombin time (PT) assay and a 1 x 5 x vize
CC   activated partial thromboplastin time (aPTT) assay, calculating
CC   prolongation of clotting effected by the isolated protein in each of
CC   the PT and aPTT assay, with respect to a baseline clotting time for
CC   each assay, where prolongation of clotting is calculated as fold
CC   elevation of clotting time relative to a baseline clotting value, where
CC   a doubling of clotting time is deemed a two-fold elevation; and
CC   calculating a PT to aPTT prolongation ratio, where a ratio at least
CC   one is indicative of factor VIIa/TF inhibitory activity. The method is
```

```
CC   useful for determining if a protein has factor VIIa/TF inhibitory
CC   activity.
XX
SQ   Sequence 461 BP; 134 A; 99 C; 115 G; 113 T; 0 other;


alignment scores:
              Quality:    6.00              Length:       6
                Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_84_92 x AAZ10439   ..

 Align seq 1/1   to: AAZ10439   from: 1   to: 461

        1 GluGluProProGluGlu 6
          ||||||||||||||||||
      169 GAGGAACCCCCTGAGGAG 186

seq name: /tmp/ifw_pdata/pcseq/pcnoseq.pep, NA:  1 - 461 (461)*3

seq documentation block:
ID   AAA73362 standard; cDNA; 461 BP.
XX
AC   AAA73362;
XX
DT   19-DEC-2000  (first entry)
XX
DE   A. caninum nematode-extracted anticoagulant protein AcaNAP5 cDNA.
XX
KW   Nematode-extracted anticoagulant protein; AcaNAP5; blood clotting;
KW   canine hookworm; thrombosis; vaccine; ss.
XX
OS   Ancylostoma caninum.
XX
FH   Key             Location/Qualifiers
FT   CDS             22..324
FT                   /*tag=  a
FT                   /product=  "pre-pro AcaNAP5"
FT   sig_peptide     22..78
FT                   /*tag=  b
FT                   /note=  "pre-AcaNAP5 peptide"
FT   sig_peptide     79..90
FT                   /*tag=  c
FT                   /note=  "pro-NAP5 peptide"
FT   mat_peptide     91..321
FT                   /*tag=  d
FT                   /product=  "mature NAP5"
XX
PN   US6087487-A.
XX
PD   11-JUL-2000.
XX
PF   12-FEB-1999;   99US-0249451.
XX
PR   17-OCT-1995;   95WO-US13231.
PR   17-APR-1997;   97US-0838945.
PR   18-OCT-1994;   94US-0326110.
PR   05-JUN-1995;   95US-0461965.
PR   05-JUN-1995;   95US-0465380.
PR   05-JUN-1995;   95US-0486397.
PR   05-JUN-1995;   95US-0486399.
XX
PA   (CORV-) CORVAS INT INC.
XX
PI   Lauwereys MJ,  Stanssens PEH,  Vespers LS,  Cleromons YCI,  Moyle M,
PI   Bergum PW,  Messens JHL,  LaRoche YR,  Vlasuk GP;
XX
DR   WPI; 2000-531459/48.
DR   P-PSDB; AAB15288.
XX
PT   New cDNA molecule encoding a protein having factor Xa inhibitory
PT   activity for preventing and treating blood clotting disorders,
PT   comprises nematode extracted anticoagulant protein domains
XX
PS   Claim 2; Fig 1; 19pp; English.
XX
CC   The present sequence is the coding sequence for the Ancyclostoma caninum
CC   nematode extracted anticoagulant protein AcaNAP5. Proteins of the kind
CC   encoded by this sequence have been shown to be effective at preventing
CC   blood clotting without causing excessive bleeding. The protein can be
CC   used in blood collection tubes to aid the isolation of plasma from the
CC   blood, to prevent thrombosis which may be linked to the rupture of an
CC   atherosclerotic plaque, acute myocardial infarction, angina,
CC   thrombolytic therapy, percutaneous transluminal coronary angioplasty,
CC   disseminated intravascular coagulopathy, infection, cancer and septic
CC   shock, and to produce antibodies. In the latter instance, the antibodies
CC   can be raised in order to detect infection by nematodes (the coding
CC   sequence given here can also be used for this) or as diagnostic tests.
CC   The protein can also be used as a vaccine against nematode parasites.
XX
SQ   Sequence 461 BP; 134 A; 99 C; 115 G; 113 T; 0 other;


alignment scores:
              Quality:    6.00              Length:       6
                Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_84_92 x AAA73362   ..

 Align seq 1/1   to: AAA73362   from: 1   to: 461

        1 GluGluProProGluGlu 6
          ||||||||||||||||||
      169 GAGGAACCCCCTGAGGAG 186

seq name: /tmp/ifw_pdata/pcseq/pcnoseq.pep, NA:1-461..AL.AAX13904

seq_documentation block:
ID   AAX13904 standard; DNA; 534 BP.
XX
AC   AAX13904;
XX
DT   19-MAR-1999  (first entry)
XX
DE   Enterococcus faecalis genome contig SEQ ID No:967.
XX
KW   Enterococcus faecalis; contig; detection; Enterococcal infection;
KW   vaccine; attenuation; computer readable medium; ds.
XX
OS   Enterococcus faecalis.
XX
PN   WO9854959-A2.
XX
PD   12-NOV-1998.
XX
PF   04-MAY-1998;   98WO-US08985.
XX
PR   14-NOV-1997;   97US-0066009.
PR   06-MAY-1997;   97US-0044031.
PR   16-MAY-1997;   97US-0046655.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Barash SC,  Dillon PJ,  Kunsch CA;
XX
DR   WPI; 1999-018171/04.
XX
PT   New isolated Enterococcus faecalis polynucleotides and polypeptides
PT   - used to develop products for the detection of Enterococcus and for
PT   use in vaccines for prevention or attenuation of Enterococcus
PT   infection.
XX
```

```
PS   Claim 1; Page 2070; 2084pp; English.
XX
CC   A computer readable medium has been developed which has recorded on it
CC   982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC   AAX12938 to AAX13919 represent these nucleotide sequences which are
CC   primary nucleotide sequences, also known as contigs. The computer based
CC   system can identify fragments of the Enterococcus faecalis genome with
CC   commercial importance. The products can be used to detect the presence
CC   of Enterococcus faecalis in samples. They can also be used for
CC   diagnosing Enterococcal infection in an animal and monitoring
CC   progression of disease, and for identifying agents which can be used to
CC   modulate the growth or pathogenicity of Enterococcus faecalis, or
CC   another related organism, in vivo or in vitro. In particular the
CC   polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC   can be used in vaccines to prevent or attenuate an Enterococcal
CC   infection.
XX
SQ   Sequence 534 BP; 200 A; 132 C; 109 G; 88 T; 5 other;


alignment_scores:
             Quality:    6.00              Length:       6
               Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_84_92 x AAX13904    .

 Align seq 1/1  to: AAX13904  from: 1  to: 534

        3 ProProGluGluProGlu 8
          ||||||||||||||||||
      399 CCGCCCGAAGAACCAGAG 416

seq_name: "/cgi2/tmp/pto_seq_fromscp_NAlCe_9A: AAV83019

seq_documentation_block:
ID   AAV83019 standard; cDNA; 537 BP.
XX
AC   AAV83019;
XX
DT   18-NOV 1998  (first entry)
XX
DE   Brugia malayi transglutaminase coding strand DNA nBmlG537.
XX
KW   Dirofilaria immitis transglutaminase; TG; parasitic nematode infection;
KW   Brugia malayi transglutaminase; Onchocerca volvulus transglutaminase; ss
XX
OS   Brugia malayi.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..537
FT                   /*tag= a
FT                   /product= "B. malayi transglutaminase PBmlG129"
FT                   /note= "CDS does not contain a stop codon"
XX
PN   WO9824887-A2.
XX
PD   11-JUN-1998.
XX
PF   02-DEC-1997;   97WO-US21999.
XX
PR   03-DEC-1996;   96US-0781420.
XX
PA   (BESK-) BESKA CORP.
XX
PI   Chandrashekar R, Mehta K;
XX
DR   WPI; 1998-333309/29.
DR   P-PSDB; AAW70272.
XX
PT   New isolated parasitic nematode transglutaminase genes - used to
PT   develop products for the detection, prevention and treatment of
PT   infections by parasitic nematodes
XX
PS   Claim 3; Pages 99-100; 129pp; English.
XX
CC   The present sequence represents a Brugia malayi transglutaminase (TG)
CC   coding strand DNA nBmlG537. The claimed complementary strand is shown
CC   in AAV83035. The invention also claims for other TG protein sequences
CC   and the corresponding DNA sequences encoding these proteins is later
CC   from B. malayi (AAV83018 and AAW70271), Dirofilaria immitis (AAV83017 and
CC   AAW70270) and Onchocerca volvulus (AAV83020 and AAW70273). The nematode
CC   TG DNA and protein sequences are claimed to be useful for the diagnosis,
CC   prevention and treatment of parasitic nematode infection.
XX
SQ   Sequence 537 BP; 179 A; 81 C; 128 G; 149 T; 0 other;


alignment_scores:
             Quality:    6.00              Length:       6
               Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_84_92 x AAV83019    .

 Align seq 1/1  to: AAV83019  from: 1  to: 537

        1 GluGluProProGluGlu 6
          |||| ||||||||||||||
      537 GAGGAACCCACCGGAAGAA 454
```

```
CM of: US-09-126-945B-2_COPY_84_92.*** Issued Patents NA *    Q* * ***    ***

Date: Sep 11, 2001, 4:53 PM

About: Ros.ds wr.tt.n by ***. Sc*** ** * ****    ***.*** 1.5
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
  -MODEL trames_p2n.model -DEV xl
  -Q=/***/***SEI_**p*3_***dl*45*p.as*_2***_***3_***11_1***/***_***_***.*** -DB
  -DB Issued_Patents_NA -QFMT fastap -SUFFIX=olip2n.rni
  -GAP*P 4.500 -GAPEXT=0.050  MINMATCH 0.100 -LocPCL 0.000
  -LocPEXT=0.000 -QGAPOP 4.500 -QGAPEXT 0.050 -XGAP*P 60.000
  -XGAPEX1 60.000 -YGAPOP 6.000 -YGAPEXT 7.000 -YGAP*P 60.000
  -YGAPEXT=60.000  DPTOP 6.000  DPLEXT=7.000 -START 1
  -MATRIX=oligo -TRANS=human40.cdi -LIST 45 -D*SALIGN 200
  -THR_SCORE quality -THR_MIN=1 -ALIGN=15 -MODE LOCAL  OUTFMT pir
  -NORM=ext  MINLEN 9  MAXLEN 2000000000
  -USER US09126945_*9CN1_1_91 -NCPU 6 -IPPU 3  LeNGLH5  NO_XLIXY
   WAIT  THREADS=1

Search information block:
Query: US-09-126-945B-2_COPY_84_92
Query length: 9
Database: Issued_Patents_NA:*
Database sequences: 324599
Database length: 94655562
Search time (sec): 822.720000

WARN: XGAP*P and YGAP*P must be equal. Assuming YGAP*P XGAP*P 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000


score_list:
Sequence              Strd Orig    ZScore   BScore Len  ! Documentation  ..
/*nn2_6_pt-data/2/ina/5B_COMB.seq:US-08-465-380-1 +     6.00   126.14   20.69     234  !
/*nn2_6_pt-data/2/ina/5B_COMB.seq:US-08-490-478-2 +     6.00   126.14   20.69     234  !
/*nn2_6_pt-data/2/ina/5B_COMB.seq:US-08-486-847-1 +     6.00   126.14   20.69     234  !
/*nn2_6_pt-data/2/ina/5B_COMB.seq:US-08-486-696-1 +     6.00   126.14   20.69     234  !
/*nn2_6_pt-data/2/ina/5B_COMB.seq:US-08-462-965-1 +     6.00   126.14   20.69     234  !
/*nn2_6_pt-data/2/ina/5B_COMB.seq:US-08-*06-*50A-2 +     6.00   126.14   20.69     234  !
/*nn2_6_pt-data/2/ina/5B_COMB.seq:US-08-*83-641-1 +     6.00   126.14   20.69     234  !
/*nn2_6_pt-data/2/ina/6A_COMB.seq:US-09-249-472-1 +     6.00   126.14   20.69     234  !
/*nn2_6_pt-data/2/ina/6A_COMB.seq:US-09-249-470-1 +     6.00   126.14   20.69     234  !
/*nn2_6_pt-data/2/ina/6A_COMB.seq:US-09-249-462-1 +     6.00   126.14   20.69     234  !
/*nn2_6_pt-data/2/ina/6A_COMB.seq:US-09-*09-*75-1 +     6.00   126.14   20.69     234  !
/*nn2_6_pt-data/2/ina/6A_COMB.seq:US-09-249-467-1 +     6.00   126.14   20.69     234  !
/*nn2_6_pt-data/2/ina/6A_COMB.seq:US-09-249-449-1 +     6.00   126.14   20.69     234  !
/*nn2_6_pt-data/2/ina/5B_COMB.seq:US-08-*41-***-2 +     6.00   122.60   20.69     408  !
/*nn2_6_pt-data/2/ina/5B_COMB.seq:US-08-465-380-2 +     6.00   121.92   20.01     461  !
/*nn2_6_pt-data/2/ina/5B_COMB.seq:US-08-480-*78-2       6.00   121.92   20.01     461  !
/*nn2_6_pt-data/2/ina/5B_COMB.seq:US-08-486-847-1 +     6.00   121.92   20.01     461  !
/*nn2_6_pt-data/2/ina/5B_COMB.seq:US-08-*06-*50A-2 +     6.00   121.92   20.01     461  !
/*nn2_6_pt-data/2/ina/5B_COMB.seq:US-08-462-965-1 +     6.00   121.92   20.01     461  !
/*nn2_6_pt-data/2/ina/5B_COMB.seq:US-08-*83-641-1 +     6.00   121.92   20.01     461  !
/*nn2_6_pt-data/2/ina/6A_COMB.seq:US-09-249-472-1 +     6.00   121.92   20.01     461  !
/*nn2_6_pt-data/2/ina/6A_COMB.seq:US-09-249-470-1 +     6.00   121.92   20.01     461  !
/*nn2_6_pt-data/2/ina/6A_COMB.seq:US-09-249-462-1 +     6.00   121.92   20.01     461  !
/*nn2_6_pt-data/2/ina/6A_COMB.seq:US-09-*09-*75-1 +     6.00   121.92   20.01     461  !
/*nn2_6_pt-data/2/ina/6A_COMB.seq:US-09-249-467-1 +     6.00   121.92   20.01     461  !
/*nn2_6_pt-data/2/ina/6A_COMB.seq:US-09-249-449-1 +     6.00   121.92   20.01     461  !
/*nn2_6_pt-data/2/ina/6A_COMB.seq:US-09-249-419-1 +     6.00   121.92   20.01     461  !
/*nn2_6_pt-data/2/ina/6A_COMB.seq:US-09-*41-***-2 +     6.00   122.60   20.69     408  !
/*nn2_6_pt-data/2/ina/5B_COMB.seq:US-08-749-527-1       6.00   108.86   189.82   3523  !
/*nn2_6_pt-data/2/ina/6A_COMB.seq:US-08-*85-420-1 +     6.00    94.47   633.11   1537*  !
/*nn2_6_pt-data/2/ina/5B_COMB.seq:US-08-*67-*24*-6 +     5.00   119.20   50.39     46  !
/*nn2_6_pt-data/2/ina/5B_COMB.seq:US-08-*30-***-1 +     5.00   114.52   91.81     75  !
/*nn2_6_pt-data/2/ina/5B_COMB.seq:US-09-072-*24-1 +     5.00   114.52   91.81     75  !
/*nn2_6_pt-data/2/ina/5B_COMB.seq:US-08-*2*-***-* +     5.00   111.**   **.**     **  !
/*nn2_6_pt-data/2/ina/6A_COMB.seq:US-09-*0*-***-** +     5.00   111.**   **.**     **  !
/*nn2_6_pt-data/2/ina/5B_COMB.seq:US-08-464-418-1 +     5.00   109.74   169.68    159  !
/*nn2_6_pt-data/2/ina/5B_COMB.seq:US-09-*61-780A-204     5.00   107.08   238.*9    241  !
/*nn2_6_pt-data/2/ina/5A_COMB.seq:US-09-*59-*84-25 +     5.00   106.33   270.27    281
/*nn2_6_pt-data/2/ina/5A_COMB.seq:US-09-449-548-25 +     5.00   106.33   270.27    281
/*nn2_6_pt-data/2/ina/5B_COMB.seq:US-08-*09-*74-25 +     5.00   106.33   270.27    281
/*nn2_6_pt-data/2/ina/5B_COMB.seq:US-08-*09-*72-25 +     5.00   106.33   270.27    281
/*nn2_6_pt-data/2/ina/6A_COMB.seq:US-08-444-818-24 +     5.00   106.09   273.94    282

>>seq**** ***.*_*/pt-data/*/ina/5B_COMB.seq:US-08-465-380-1

seq_documentation_block:
; Sequence 1, Application US/08465380
; Patent No. 5668894
;  GENERAL INFORMATION:
;     APPLICANT:  George P. Vlasuk, Patric H. Stanssens,
;     APPLICANT:  Joris H.L. Mensens, Marc J. Lauwereys,
;     APPLICANT:  Yves R. Laroche, Laurent S. Jospers,
;     APPLICANT:  Yannick G.J. Gansemans, Matthew Moyle,
;     APPLICANT:  Peter W. Bernam
;     TITLE OF INVENTION:  NEMATODE EXTRACTED ANTICOAGULANT
;     TITLE OF INVENTION:  PROTEIN
;     NUMBER OF SEQUENCES:  456
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Lyon & Lyon
;       STREET:  633 West Fifth Street
;       STREET:  Suite 4700
;       CITY:  Los Angeles
;       STATE:  California
;       COUNTRY:  U.S.A.
;       ZIP:  90071
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5" Diskette, 1.44 MB
;       MEDIUM TYPE:  storage
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  IBM P.C. DOS 5.0
;       SOFTWARE:  Word Perfect 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/465,480
;       FILING DATE:  June 5, 1995
;       CLASSIFICATION:  530
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/326,110
;       FILING DATE:  October 28, 1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Boggs, SUZANNE L.
;       REGISTRATION NUMBER:  30,158
;       REFERENCE/DOCKET NUMBER:  2*.368
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (213) 489-1600
;       TELEFAX:  (213) 955-0440
;       TELEX:  67-3510
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  234 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-465-380-1


alignment_scores:
                quality:    6.00              Length:      6
                  Ratio:    1.**                Gaps:      0
 Percent similarity:  ***.**    Percent identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_84_92 x US-08-465-380-1    ..

 Query:    1  ***********  234

          SignalProfileValue 6
          ||||||||||||||||||
Db:    234  CAGGAG  96
```

```
Seq name: Copy of pta-data/US_Grp_08 Mon Sep 17 09:58:26 200 2

Seq documentation block:
; Sequence 2, Application US/08480478
; Patent No. 5864009
;  GENERAL INFORMATION:
;     APPLICANT:  GEORGE P. VLASUK; PATRICK ERAT
;     APPLICANT:  HUGO STANSSENS; JORIS HILDA
;     APPLICANT:  LIEVEN MESSENS; MARC JOZEF
;     APPLICANT:  LAUWEREYS; YVES RENE LAROCHE;
;     APPLICANT:  LAURENT STEPHANE JESPERS; and
;     APPLICANT:  YANNICK GEORGES JOZEF
;     APPLICANT:  GANSEMANS
;     TITLE OF INVENTION:  NEMATODE-EXTRACTED ANTI
;     TITLE OF INVENTION:  COAGULANT PROTEIN
;     NUMBER OF SEQUENCES:  86
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Lyon & Lyon
;       STREET:  633 West Fifth Street
;       STREET:  Suite 4700
;       CITY:  Los Angeles
;       STATE:  California
;       COUNTRY:  U.S.A.
;       ZIP:  90071
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;       MEDIUM TYPE:  storage
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  IBM P.C. DOS 5.0
;       SOFTWARE:  FastSEQ Version 1.5
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/480,478
;       FILING DATE:  06-JUN 1995
;       CLASSIFICATION:  530
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/426,110
;       FILING DATE:  18 OCTOBER 1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BIGGS, SUZANNE L.
;       REGISTRATION NUMBER:  30,158
;       REFERENCE/DOCKET NUMBER:  208/290
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (213) 489 1600
;       TELEFAX:  (213) 955 0440
;       TELEX:  67 3510
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  234 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US 08 480 478 2


Alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US 09 126 945B 2 COPY 84 92 x US 08 480 478 2    ..

 Align Seq 1/1   to: US 08 480 478 2  from: 1  to: 234

        1 GluGluProProGluGlu 6
          ||||||||||||||||||
       79 GAGGAACCCCCTGAGGAG 96

Seq name: Copy of pta-data/US_Grp_08 SAM Sep 17 09:58:26 200 1

Seq documentation block:
; Sequence 1, Application US/08486397
; Patent No. 5864009
;  GENERAL INFORMATION:
;     APPLICANT:  George P. Vlasuk, Patric H. Stanssens,
;     APPLICANT:  Joris H.L. Messens, Marc J. Lauwereys,
;     APPLICANT:  Yves R. Laroche, Laurent S. Jespers,
;     APPLICANT:  Yannick G.J. Gansemans, Matthew Moyle,
;     APPLICANT:  Peter W. Berquin
;     TITLE OF INVENTION:  NEMATODE EXTRACTED ANTI-COAGULANT
;     TITLE OF INVENTION:  PROTEIN
;     NUMBER OF SEQUENCES:  357
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Lyon & Lyon
;       STREET:  633 West Fifth Street
;       STREET:  Suite 4700
;       CITY:  Los Angeles
;       STATE:  California
;       COUNTRY:  U.S.A.
;       ZIP:  90071
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;       MEDIUM TYPE:  storage
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  IBM P.C. DOS 5.0
;       SOFTWARE:  Word Perfect 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/486,397
;       FILING DATE:  June 5, 1995
;       CLASSIFICATION:  530
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/426,110
;       FILING DATE:  October 18, 1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BIGGS, SUZANNE L.
;       REGISTRATION NUMBER:  30,158
;       REFERENCE/DOCKET NUMBER:  213/269
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (213) 489-1600
;       TELEFAX:  (213) 955 0440
;       TELEX:  67 3510
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  234 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US 08 486 397 1


Alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment Block:
 US 09-126 945B-2 COPY 84 92 x US 08 486 397 1    ..

 Align Seq 1/1   to: US 08 486 397 1  from: 1  to: 234

        1 GluGluProProGluGlu 6
          ||||||||||||||||||
       79 GAGGAACCCCCTGAGGAG 96

Seq name: Copy of pta-data/US_Grp_08 COMB.seq:US 08 486 399 1

Seq documentation block:
; Sequence 1, Application US/08486399
; Patent No. 5864344
;  GENERAL INFORMATION:
;     APPLICANT:  George P. Vlasuk, Patric H. Stanssens,
;     APPLICANT:  Joris H.L. Messens, Marc J. Lauwereys,
;     APPLICANT:  Yves R. Laroche, Laurent S. Jespers,
;     APPLICANT:  Yannick G.J. Gansemans, Matthew Moyle,
;     APPLICANT:  Peter W. Berquin
```

```
;     TITLE OF INVENTION:  NEMATODE EXTRACTED ANTICOAGULANT
;     TITLE OF INVENTION:  PROTEIN
;     NUMBER OF SEQUENCES:  456
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Lyon & Lyon
;       STREET:  633 West Fifth Street
;       STREET:  Suite 4700
;       CITY:  Los Angeles
;       STATE:  California
;       COUNTRY:  U.S.A.
;       ZIP:  90071
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;       MEDIUM TYPE:  storage
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  IBM P.C. DOS 5.0
;       SOFTWARE:  Word Perfect 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/486,399
;       FILING DATE:  June 5, 1995
;       CLASSIFICATION:  530
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/326,110
;       FILING DATE:  October 18, 1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BIGGS, SUZANNE L.
;       REGISTRATION NUMBER:  30,158
;       REFERENCE/DOCKET NUMBER:  214/270
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (213) 489-1600
;       TELEFAX:  (213) 955-0440
;       TELEX:  67-3510
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  234 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-486-399-1


Alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment_block:
 US-09-126-945B-2_COPY_84_92 x US-08-486-399-1    .

 Align seq 1/1  to: US-08-486-399-1  from: 1  to: 234

        1 GluGluProProGluGlu 6
          ||||||||||||||||||
       79 GAGGAACCCCCTGAGGAG 96

seq name: /tmp2_6/pto-data/2/tmp_58_2.MB_seq_1:_08-461-965_1

seq_documentation_block:
; Sequence 1, Application US/08461965
; Patent No. 5872098
;  GENERAL INFORMATION:
;     APPLICANT:  George P. Vlasuk, Patric H. Thuronyi,
;     APPLICANT:  Joris H.L. Mensens, Marc J. Lauwereys,
;     APPLICANT:  Yves R. Laroche, Laurent S. Jespers,
;     APPLICANT:  Yannick G.J. Gansemans, Matthew Moyle,
;     APPLICANT:  Peter W. Berqum
;     TITLE OF INVENTION:  NEMATODE-EXTRACTED ANTICOAGULANT
;     TITLE OF INVENTION:  PROTEIN
;     NUMBER OF SEQUENCES:  456
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Lyon & Lyon
;       STREET:  633 West Fifth Street
;       STREET:  Suite 4700
;       CITY:  Los Angeles
;       STATE:  California
;       COUNTRY:  U.S.A.
;       ZIP:  90071
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;       MEDIUM TYPE:  storage
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  IBM P.C. DOS 5.0
;       SOFTWARE:  Word Perfect 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/461,965
;       FILING DATE:  June 5, 1995
;       CLASSIFICATION:  530
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/326,110
;       FILING DATE:  October 18, 1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BIGGS, SUZANNE L.
;       REGISTRATION NUMBER:  30,158
;       REFERENCE/DOCKET NUMBER:  210/243
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (213) 489-1600
;       TELEFAX:  (213) 955-0440
;       TELEX:  67-3510
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  234 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-461-965-1


Alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment_block:
 US-09-126-945B-2_COPY_84_92 x US-08-461-965-1    .

 Align seq 1/1  to: US-08-461-965-1  from: 1  to: 234

        1 GluGluProProGluGlu 6
          ||||||||||||||||||
       79 GAGGAACCCCCTGAGGAG 96

seq name: /tmp_6/pto-data/2/tmp_58_2.MB_seq_1:_08-461-965_1

seq_documentation_block:
; Sequence 2, Application US/08326110A
; Patent No. 5945275
;  GENERAL INFORMATION:
;     APPLICANT:  GEORGE P. VLASUK; PATRICK ERIC
;     APPLICANT:  BIGGS, SUZANNE L.; JESPERS, HILDA
;     APPLICANT:  LIEVEN MESSENS; MARC JOZEF
;     APPLICANT:  LAUWEREYS; YVES RENE LAROCHE;
;     APPLICANT:  LAUREN; STEPHANE JESPERS; and
;     APPLICANT:  YANNICK GEORGES JOZEF
;     APPLICANT:  GANSEMANS
;     TITLE OF INVENTION:  NEMATODE-EXTRACTED ANTI
;     TITLE OF INVENTION:  COAGULANT PROTEIN
;     NUMBER OF SEQUENCES:  86
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Lyon & Lyon
;       STREET:  633 West Fifth Street
;       STREET:  Suite 4700
;       CITY:  Los Angeles
;       STATE:  California
;       COUNTRY:  U.S.A.
;       ZIP:  90071
;     COMPUTER READABLE FORM:
```

```
;       MEDIUM TYPE:  4.5" Diskette, 1.44 Mb
;       MEDIUM TYPE:  Storage
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  IBM P.C. DOS 5.0
;       SOFTWARE:  FastSEQ Version 1.5
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/326,110A
;       FILING DATE:  18 OCTOBER 1994
;       CLASSIFICATION:  530
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BIGGS, SUZANNE L.
;       REGISTRATION NUMBER:  30,158
;       REFERENCE/DOCKET NUMBER:  298/290
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (213) 489-1600
;       TELEFAX:  (213) 955 0440
;       TELEX:  67 3510
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  234 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-326-110A-2


Alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-126-945B-2_COPY_84_92 x US-08-326-110A-2  ..

 Align seq 1/1  to: US-08-326-110A-2  from: 1  to: 234

       1 GluGluProProGluGlu 6
         ||||||||||||||||||
      79 GAGGAACCCCCTGAGGAG 96

Seq name: /cgn, 6/ptodata/2/fina/US-09B-seq-0--or-ext-631.

Seq documentation block:
; Sequence 1, Application US/08644641
; Patent No. 5955294
;  GENERAL INFORMATION:
;     APPLICANT:  Vlasuk, George P. Vlasuk
;     APPLICANT:  Stanssens, Patrick Eric Bajo
;     APPLICANT:  Mennens, Joris Hilda Lieven
;     APPLICANT:  Lauwereys, Marc Josef
;     APPLICANT:  Laroche, Yves Rene
;     APPLICANT:  Jespers, Laurent Stephane
;     APPLICANT:  Gansemans, Yannick Georges Jozef
;     APPLICANT:  Moyle, Matthew
;     APPLICANT:  Bergum, Peter W.
;     TITLE OF INVENTION:  NEMATODE EXTRACTED ANTICOAGULANT
;     TITLE OF INVENTION:  PROTEIN
;     NUMBER OF SEQUENCES:  456
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Lyon & Lyon
;       STREET:  633 West Fifth Street
;       STREET:  Suite 4700
;       CITY:  Los Angeles
;       STATE:  California
;       COUNTRY:  U.S.A.
;       ZIP:  90071
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  4.5" Diskette, 1.44 Mb
;       MEDIUM TYPE:  Storage
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  IBM P.C. DOS 5.0
;       SOFTWARE:  Word Perfect 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/634,641
;       FILING DATE:  April 19, 1996
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/IB94/01231
;       FILING DATE:  October 17, 1995
;       APPLICATION NUMBER:  08/486,399
;       FILING DATE:  June 5, 1995
;       APPLICATION NUMBER:  08/486,397
;       FILING DATE:  June 5, 1995
;       APPLICATION NUMBER:  08/465,380
;       FILING DATE:  June 5, 1995
;       APPLICATION NUMBER:  08/461,965
;       FILING DATE:  June 5, 1995
;       APPLICATION NUMBER:  08/326,110
;       FILING DATE:  October 18, 1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BIGGS, SUZANNE L.
;       REGISTRATION NUMBER:  30,158
;       REFERENCE/DOCKET NUMBER:  219/136
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (213) 489-1600
;       TELEFAX:  (213) 955 0440
;       TELEX:  67 3510
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  234 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-634-641-1


Alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-126-945B-2_COPY_84_92 x US-08-634-641-1  ..

 Align seq 1/1  to: US-08-634-641-1  from: 1  to: 234

       1 GluGluProProGluGlu 6
         ||||||||||||||||||
      79 GAGGAACCCCCTGAGGAG 96

Seq name: /cgn, 6/ptodata/2/fina/US-09A-PCT-MB-seq-US-09-249-471-1.

Seq documentation block:
; Sequence 1, Application US/09249471
; Patent No. 6040441
;  GENERAL INFORMATION:
;     APPLICANT:  Vlasuk, George Phillip
;     APPLICANT:  Stanssens, Patrick Eric Bajo
;     APPLICANT:  Mennens, Joris Hilda Lieven
;     APPLICANT:  Lauwereys, Marc Josef
;     APPLICANT:  Laroche, Yves Rene
;     APPLICANT:  Jespers, Laurent Stephane
;     APPLICANT:  Gansemans, Yannick Georges Jozef
;     APPLICANT:  Moyle, Matthew
;     APPLICANT:  Bergum, Peter W.
;     TITLE OF INVENTION:  NEMATODE EXTRACTED SERINE PROTEASE
;     TITLE OF INVENTION:  INHIBITORS AND ANTICOAGULANT
;     TITLE OF INVENTION:  PROTEIN
;     NUMBER OF SEQUENCES:  456
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Lyon & Lyon
;       STREET:  633 West Fifth Street
;       STREET:  Suite 4700
;       CITY:  Los Angeles
```

```
;       STATE:  California
;       COUNTRY:  U.S.A.
;       ZIP:  90071
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;       MEDIUM TYPE:  storage
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  IBM P.C. DOS 5.0
;       SOFTWARE:  Word Perfect 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/249,472
;       FILING DATE:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/809,455
;       FILING DATE:  April 17, 1997
;       APPLICATION NUMBER:  PCT/US95/13231
;       FILING DATE:  October 17, 1995
;       APPLICATION NUMBER:  08/486,399
;       FILING DATE:  June 5, 1995
;       APPLICATION NUMBER:  08/486,397
;       FILING DATE:  June 5, 1995
;       APPLICATION NUMBER:  08/465,880
;       FILING DATE:  June 5, 1995
;       APPLICATION NUMBER:  08/461,965
;       FILING DATE:  June 5, 1995
;       APPLICATION NUMBER:  08/326,110
;       FILING DATE:  October 18, 1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BIGGS, SUZANNE L.
;       REGISTRATION NUMBER:  30,158
;       REFERENCE/DOCKET NUMBER:  216/270
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (213) 489-1600
;       TELEFAX:  (213) 955-0440
;       TELEX:  67-3510
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  234 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-09-249-471-1


alignment_scores:
             Quality:    6.00              Length:       6
               Ratio:   1.000              Gaps:         0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_84_92 x US-09-249-471-1    ..

 Align seq 1/1  to: US-09-249-471-1  from: 1  to: 234

        1 GluGluProProGluGlu 6
          ||||||||||||||||||
       79 GAGGAACCCCCTGAGGAG 96

seq_name: /cgn2_6/ptodata/2/iia/A_COMB_seq_US_09_249_472_1

seq_documentation_block:
; Sequence 1, Application US/09249472
; Patent No. 6046318
;  GENERAL INFORMATION:
;     APPLICANT:  Vlasuk, George Phillip
;     APPLICANT:  Stanssens, Patrick Eric Hugo
;     APPLICANT:  Messens, Joris Hilda Lieven
;     APPLICANT:  Lauwereys, Marc Josef
;     APPLICANT:  Laroche, Yves Rene
;     APPLICANT:  Jespers, Laurent Stephane
;     APPLICANT:  Gansemans, Yannick Georges Hubert
;     APPLICANT:  Moyle, Matthew
;     APPLICANT:  Bergum, Peter W.
;     TITLE OF INVENTION:  NEMATODE EXTRACTED SERINE PROTEASE
;     TITLE OF INVENTION:  INHIBITORS AND ANTICOAGULANT
;     TITLE OF INVENTION:  PROTEIN
;     NUMBER OF SEQUENCES:  456
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Lyon & Lyon
;       STREET:  633 West Fifth Street
;       STREET:  Suite 4700
;       CITY:  Los Angeles
;       STATE:  California
;       COUNTRY:  U.S.A.
;       ZIP:  90071
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;       MEDIUM TYPE:  storage
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  IBM P.C. DOS 5.0
;       SOFTWARE:  Word Perfect 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/249,472
;       FILING DATE:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/809,455
;       FILING DATE:  April 17, 1997
;       APPLICATION NUMBER:  PCT/US95/13231
;       FILING DATE:  October 17, 1995
;       APPLICATION NUMBER:  08/486,399
;       FILING DATE:  June 5, 1995
;       APPLICATION NUMBER:  08/486,397
;       FILING DATE:  June 5, 1995
;       APPLICATION NUMBER:  08/465,880
;       FILING DATE:  June 5, 1995
;       APPLICATION NUMBER:  08/461,965
;       FILING DATE:  June 5, 1995
;       APPLICATION NUMBER:  08/326,110
;       FILING DATE:  October 18, 1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BIGGS, SUZANNE L.
;       REGISTRATION NUMBER:  30,158
;       REFERENCE/DOCKET NUMBER:  216/270
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (213) 489-1600
;       TELEFAX:  (213) 955-0440
;       TELEX:  67-3510
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  234 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-09-249-472-1


alignment_scores:
             Quality:    6.00              Length:       6
               Ratio:   1.000              Gaps:         0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_84_92 x US-09-249-472-1    ..

 Align seq 1/1  to: US-09-249-472-1  from: 1  to: 234

        1 GluGluProProGluGlu 6
          ||||||||||||||||||
       79 GAGGAACCCCCTGAGGAG 96

seq_name: /cgn2_6/ptodata/2/iia/A_COMB_seq_US_09_249_471_1

seq_documentation_block:
; Sequence 1, Application US/09249451
; Patent No. 6087487
;  GENERAL INFORMATION:
```

```
;     APPLICANT:  Vlasuk, George Phillip
;     APPLICANT:  Stanssens, Patrick Eric Hugo
;     APPLICANT:  Mossens, Joris Hilda Lieven
;     APPLICANT:  Lauwereys, Marc Josef
;     APPLICANT:  Laroche, Yves Rene
;     APPLICANT:  Jespers, Laurent Stephane
;     APPLICANT:  Ganseman, Yannick Georges Jozef
;     APPLICANT:  Moyle, Matthew
;     APPLICANT:  Bergum, Peter W.
;     TITLE OF INVENTION:  NEMATODE EXTRACTED SERINE PROTEASE
;     TITLE OF INVENTION:  INHIBITORS AND ANTICOAGULANT
;     TITLE OF INVENTION:  PROTEIN
;     NUMBER OF SEQUENCES:  456
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Lyon & Lyon
;       STREET:  633 West Fifth Street
;       STREET:  Suite 4700
;       CITY:  Los Angeles
;       STATE:  California
;       COUNTRY:  U.S.A.
;       ZIP:  90071
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;       MEDIUM TYPE:  storage
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  IBM P.C. DOS 5.0
;       SOFTWARE:  Word Perfect 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/249,451
;       FILING DATE:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/809,455
;       FILING DATE:  April 17, 1997
;       APPLICATION NUMBER:  PCT/US95/13231
;       FILING DATE:  October 17, 1995
;       APPLICATION NUMBER:  08/486,399
;       FILING DATE:  June 5, 1995
;       APPLICATION NUMBER:  08/486,397
;       FILING DATE:  June 5, 1995
;       APPLICATION NUMBER:  08/465,880
;       FILING DATE:  June 5, 1995
;       APPLICATION NUMBER:  08/461,965
;       FILING DATE:  June 5, 1995
;       APPLICATION NUMBER:  08/326,110
;       FILING DATE:  October 18, 1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BIGGS, SUZANNE L.
;       REGISTRATION NUMBER:  40,158
;       REFERENCE/DOCKET NUMBER:  216/270
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (214) 489 1600
;       TELEFAX:  (214) 955 0440
;       TELEX:  67 4510
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  234 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US 09 249 451 1


alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US 09 126 945B 2 COPY 84_92 x US 09 249 451 1     .

 Align seq 1/1  to: US 09 249 451 1  from: 1  to: 234

       1 GluGluProProGluGlu 6
```

```
          ctc ctcttttttttt
       79 GAGGAACCCCCTGAGGAG 96

<seq_doc  /seq_doc_type="pct-doc"/seq_doc_GAME="seq.US 08 809 455 1

seq documentation block:
; Sequence 1, Application US/08809455
; Patent No. 6008026
;  GENERAL INFORMATION:
;     APPLICANT:  Vlasuk, George Phillip
;     APPLICANT:  Stanssens, Patrick Eric Hugo
;     APPLICANT:  Mossens, Joris Hilda Lieven
;     APPLICANT:  Lauwereys, Marc Josef
;     APPLICANT:  Laroche, Yves Rene
;     APPLICANT:  Jespers, Laurent Stephane
;     APPLICANT:  Ganseman, Yannick Georges Jozef
;     APPLICANT:  Moyle, Matthew
;     APPLICANT:  Bergum, Peter W.
;     TITLE OF INVENTION:  NEMATODE EXTRACTED SERINE PROTEASE
;     TITLE OF INVENTION:  INHIBITORS AND ANTICOAGULANT
;     TITLE OF INVENTION:  PROTEIN
;     NUMBER OF SEQUENCES:  456
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Lyon & Lyon
;       STREET:  633 West Fifth Street
;       STREET:  Suite 4700
;       CITY:  Los Angeles
;       STATE:  California
;       COUNTRY:  U.S.A.
;       ZIP:  90071
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;       MEDIUM TYPE:  storage
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  IBM P.C. DOS 5.0
;       SOFTWARE:  Word Perfect 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/809,455
;       FILING DATE:  April 17, 1997
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/13231
;       FILING DATE:  October 17, 1995
;       APPLICATION NUMBER:  08/486,399
;       FILING DATE:  June 5, 1995
;       APPLICATION NUMBER:  08/486,397
;       FILING DATE:  June 5, 1995
;       APPLICATION NUMBER:  08/465,880
;       FILING DATE:  June 5, 1995
;       APPLICATION NUMBER:  08/461,965
;       FILING DATE:  June 5, 1995
;       APPLICATION NUMBER:  08/326,110
;       FILING DATE:  October 18, 1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BIGGS, SUZANNE L.
;       REGISTRATION NUMBER:  40,158
;       REFERENCE/DOCKET NUMBER:  216/270
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (214) 489 1600
;       TELEFAX:  (214) 955 0440
;       TELEX:  67 4510
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  234 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US 08 809 455 1


alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000
```

```
alignment_block:
 US-09-126-945B-2_COPY_84_92 x US-08-809-455-1    .

 Align seq 1/1  to: US-08-809-455-1  from: 1  to: 234

        1 GluGluProProGluGlu 6
          ||||||||||||||||||
       79 GAGGAACCCCCTGAGGAG 96

seq name: /cgn2_6/ptodata/1/iia/PA_COMB.seq US-09-249-461-1

seq_documentation_block:
; Sequence 1, Application US/09249461
; Patent No. 6096877
;  GENERAL INFORMATION:
;    APPLICANT:  Vlasuk, George Phillip
;    APPLICANT:  Stanssens, Patrick Eric Hugo
;    APPLICANT:  Messens, Joris Hilda Lieven
;    APPLICANT:  Lauwereys, Marc Josef
;    APPLICANT:  Laroche, Yves Rene
;    APPLICANT:  Jespers, Laurent Stephane
;    APPLICANT:  Gansemans, Yannick Georges Jozef
;    APPLICANT:  Moyle, Matthew
;    APPLICANT:  Berqum, Peter W.
;    TITLE OF INVENTION:  NEMATODE-EXTRACTED SERINE PROTEASE
;    TITLE OF INVENTION:  INHIBITORS AND ANTICOAGULANT
;    TITLE OF INVENTION:  PROTEIN
;    NUMBER OF SEQUENCES:  456
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  Word Perfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/249,461
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/809,455
;      FILING DATE:  April 17, 1997
;      APPLICATION NUMBER:  PCT/US95/13231
;      FILING DATE:  October 17, 1995
;      APPLICATION NUMBER:  08/486,399
;      FILING DATE:  June 5, 1995
;      APPLICATION NUMBER:  08/486,397
;      FILING DATE:  June 5, 1995
;      APPLICATION NUMBER:  08/465,380
;      FILING DATE:  June 5, 1995
;      APPLICATION NUMBER:  08/461,965
;      FILING DATE:  June 5, 1995
;      APPLICATION NUMBER:  08/326,110
;      FILING DATE:  October 18, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BIGGS, SUZANNE L.
;      REGISTRATION NUMBER:  30,158
;      REFERENCE/DOCKET NUMBER:  211/278
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  234 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-249-461 1


alignment_scores:
            Quality:     6.00              Length:       6
              Ratio:    1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment:
 US-09-126-945B-2_COPY_84_92 x US-09-249-461-1    .

 Align seq 1/1  to: US-09-249-461-1  from: 1  to: 234

        1 GluGluProProGluGlu 6
          ||||||||||||||||||
       79 GAGGAACCCCCTGAGGAG 96

seq name: /cgn2_6/ptodata/1/iia/PA_COMB.seq US-09-249-448-1

seq_documentation_block:
; Sequence 1, Application US/09249448
; Patent No. 6121435
;  GENERAL INFORMATION:
;    APPLICANT:  Vlasuk, George Phillip
;    APPLICANT:  Stanssens, Patrick Eric Hugo
;    APPLICANT:  Messens, Joris Hilda Lieven
;    APPLICANT:  Lauwereys, Marc Josef
;    APPLICANT:  Laroche, Yves Rene
;    APPLICANT:  Jespers, Laurent Stephane
;    APPLICANT:  Gansemans, Yannick Georges Jozef
;    APPLICANT:  Moyle, Matthew
;    APPLICANT:  Berqum, Peter W.
;    TITLE OF INVENTION:  NEMATODE EXTRACTED SERINE PROTEASE
;    TITLE OF INVENTION:  INHIBITORS AND ANTICOAGULANT
;    TITLE OF INVENTION:  PROTEIN
;    NUMBER OF SEQUENCES:  456
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  Word Perfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/249,448
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/809,455
;      FILING DATE:  April 17, 1997
;      APPLICATION NUMBER:  PCT/US95/13231
;      FILING DATE:  October 17, 1995
;      APPLICATION NUMBER:  08/486,399
;      FILING DATE:  June 5, 1995
;      APPLICATION NUMBER:  08/486,397
;      FILING DATE:  June 5, 1995
;      APPLICATION NUMBER:  08/465,380
;      FILING DATE:  June 5, 1995
;      APPLICATION NUMBER:  08/461,965
;      FILING DATE:  June 5, 1995
;      APPLICATION NUMBER:  08/326,110
;      FILING DATE:  October 18, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BIGGS, SUZANNE L.
;      REGISTRATION NUMBER:  30,158
```

```
;       REFERENCE/DOCKET NUMBER:  215/270
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (213) 489 1600
;       TELEFAX:  (213) 955 0440
;       TELEX:  67 3510
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  244 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-09-249-448-1


Alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-126-945B-2-COPY-84-92 x US-09-249-448-1    .

 Align seq 1/1  to: US-09-249-448-1  from: 1  to: 244

        1 GluGluProProGluGlu 6
          ||||||||||||||||||
       79 GAGGAACCCCCTGAGGAG 96

Seq_name: /cgn2_6/ptodata/2/ina/US_09MB.seq:::US-09-981-663-69

Seq documentation block:
; Sequence 69, Application US/09981663
; Patent No. 5952176
;  GENERAL INFORMATION:
;     APPLICANT:  MCCARTHY, THOMAS VALENTINE
;     APPLICANT:  VAUGHAN, PATRICK MARTIN
;     TITLE OF INVENTION:  GLYCOSYLASE MEDIATED DETECTION OF
;     TITLE OF INVENTION:  NUCLEOTIDE SEQUENCES AT CANDIDATE LOCI
;     NUMBER OF SEQUENCES:  69
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Birch, Stewart, Kolasch & Birch, LLP
;       STREET:  P.O. Box 747
;       CITY:  Falls Church
;       STATE:  Virginia
;       COUNTRY:  USA
;       ZIP:  22040-0747
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/981,663
;       FILING DATE:  08-JAN-1998
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  SVENSSON, LEONARD F.
;       REGISTRATION NUMBER:  30,330
;       REFERENCE/DOCKET NUMBER:  1377-1200
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 205-8000
;       TELEFAX:  (703) 205-8050
;       TELEX:  248345
;   INFORMATION FOR SEQ ID NO:  69:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  408 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Homo sapiens
US-08-981-663-69


Alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-126-945B-2-COPY-84-92 x US-08-981-663-69    .

 Align seq 1/1  to: US-08-981-663-69  from: 1  to: 408

        1 GluGluProProGluGlu 6
          ||||||||||||||||||
      229 GAGGAACCCCCTGAGGAG 246

Seq_name: /cgn2_6/ptodata/2/ina/US_09MB.seq:::US-08-465-380-3

Seq documentation block:
; Sequence 3, Application US/08465380
; Patent No. 5863894
;  GENERAL INFORMATION:
;     APPLICANT:  George P. Vlasuk, Patric R. Stanssens,
;     APPLICANT:  Joris H.L. Mensens, Marc J. Lauwereys,
;     APPLICANT:  Yves R. Laroche, Laurent S. Jespers,
;     APPLICANT:  Yannick G.J. Gansemans, Matthew Moyle,
;     APPLICANT:  Peter W. Bergum
;     TITLE OF INVENTION:  NEMATODE EXTRACTED ANTICOAGULANT
;     TITLE OF INVENTION:  PROTEIN
;     NUMBER OF SEQUENCES:  356
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Lyon & Lyon
;       STREET:  633 West Fifth Street
;       STREET:  Suite 4700
;       CITY:  Los Angeles
;       STATE:  California
;       COUNTRY:  U.S.A.
;       ZIP:  90071
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;       MEDIUM TYPE:  storage
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  IBM P.C. DOS 5.0
;       SOFTWARE:  Word Perfect 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/465,380
;       FILING DATE:  June 5, 1995
;       CLASSIFICATION:  530
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/326,110
;       FILING DATE:  October 18, 1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BIGGS, SUZANNE L.
;       REGISTRATION NUMBER:  30,158
;       REFERENCE/DOCKET NUMBER:  213/268
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (213) 489 1600
;       TELEFAX:  (213) 955-0440
;       TELEX:  67 3510
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  461 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     ORIGINAL SOURCE:
;       ORGANISM:  Ancylostoma caninum
;     FEATURE:
;       NAME/KEY:  Coding Sequence
;       LOCATION:  22...321
US-08-465-380-3
```

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_84_92 x US-08-465-380-3 ..

Align Seq 1/1 to: US-08-465-380-3 from: 1 to: 471

1 GluGluProProGluGlu 6
|||||||||||||||||||
169 GAGGAACCCCCTGAGGAG 186

```
M of: US-09-126-945B-2_COPY_84_92.fas* vs EST:*  ept_format : fas

Date: Sep 14, 2001  7:17 PM

About: Results were produced by the GenCore software, version 4.5
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL frame=_p2n.model -DEV xlp
 -Q "/cgn2_7/SEQ_sp_1/US-09-126-945B-2_COPY_84_92.fas" -DB EST -QFMT fastap -SUFFIX=olip2n.rst -GAP=P 4.500
 -DB EST -QFMT=fastap -SUFFIX=olip2n.rst -GAP=P 4.500
 -GAPEXT 0.050 -MINMATCH 0.100 -LOOPCL 0.000 -LOOPEXT 0.000
 -QGAP=P 4.500 -QGAPEXT=0.050 -XGAP=P=60.000 -XGAPEXT 60.000
 -YGAP=P 6.000 -YGAPEXT=7.000 -YGAP=P=60.000 -YGAPEXT 60.000
 -DELOP=6.000 -DELEXT=7.000 -START 1 -MATRIX oligo
 -TRANS human40.cdi -LIST=45 -DOCALIGN 200 -THR_SCORE quality
 -THR_MIN=1 -ALIGN 15 -MODE LOCAL -OUTFMT pts  NORM ext  MINLEN 0
 -MAXLEN 2000000000 -USER=US09126945_@CGN1_1_4645 -NCPU=6
 -ICPU 3 -LONGLOG -NO_XLPXY -WAIT -THREADS 1

Search information block:
Query: US-09-126-945B-2_COPY_84_92
Query length: 9
Database: EST:*
Database sequences: 10228115
Database length: 4314594 54
Search time (sec): 9087.640000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT 60.000


score_list:
Sequence              Strd Orig    ZScore    EScore Len   ! Documentation ..
qb_est80:BE871340      +     9.00   171.74     1.90     945  ! BE871340 601449457F1 NIH_MGC_65
qb_est97:BG179568      +     9.00   171.43     1.98     991  ! BG179568 602428182F1 NIH_MGC_91
qb_qss28:AZ511504      +     8.00   157.10    12.44     509  ! AZ511504 IM035N19F Mouse 10kb
qb_qss26:AZ490410      -     8.00   156.47    13.18     560  ! AZ490410 IM015E16R Mouse 10kb
qb_est80:BE844192      -     8.00   156.27    13.82     577  ! BE844192 RC3_TN0094_110800_021
qb_est80:BE844191      -     8.00   156.20    13.95     583  ! BE844191 RC3_TN0094_110800_021
qb_qss9:AQ272338       -     8.00   156.14    14.06     588  ! AQ272338 qbxb0027b1r cDNA Rice
qb_est102:BG598416           7.00   147.86    49.67     122  ! BG598416 EST503216 cDNA Solanum
qb_est70:BE083507      -     7.00   141.61    90.70     812  ! BE083507 RC5_BT0645_110400_013
qb_est78:BE707851      -     7.00   141.17    95.89     833  ! BE707851 QV3_BT0544_290700_268
qb_est110:W63245       +     7.00   140.47   104.91     870  ! W63245 EST52y68b01.r1 TgME49_B
qb_est106:N60773       +     7.00   140.04   110.91     895  ! N60773 EST72y21d08.r1 TgRH_Tac
qb_est102:BG554981     +     7.00   140.00   111.42     897  ! BG554981 dac84t08.y1 NICHD_XGC
qb_est93:BF868534      +     7.00   139.75   115.00     912  ! BF868534 IL3_ET0115_121000_284
qb_est97:BG165837      +     7.00   139.67   116.20     917  ! BG165837 602444435F1 NIH_MGC_89
qb_qss5:FR0004894            7.00   139.36   120.94     937  ! Z88704 F.rubripes GSS sequence,
qb_est84:BF188465      -     7.00   139.33   121.41     939  ! BF188465 EST444752 potato stolo
qb_est74:BE141444      +     7.00   139.32   121.65     940  ! BF441444 925006A03.y1 C. reinha
qb_est106:N82633       +     7.00   139.29   122.12     942  ! N82633 EST72y38c08.r1 TgRH_Tac
qb_est43:AW135468      -     7.00   139.17   124.01     950  ! AW135468 UI-H-B11-aco-d-04-0-UI
qb_est19:AI473606      -     7.00   139.11   124.95     951  ! AI473606 qc54b03.x1 NCI_CGAP_K2
qb_est93:BF868629      +     7.00   139.02   126.46     960  ! BF868629 IL3-ET0115_121000-290-
qb_est15:AI041729      -     7.00   138.95   127.54     965  ! AI041729 qc51e06.x1 Soares_NSF
qb_est81:BE924206      -     7.00   138.85   129.17     972  ! BE924206 EST427975 potato leave
qb_qss4:CNS03P8N       -     7.00   138.37   137.30     507  ! AL254336 Tetraodon nigroviridis
qb_est101:BG458009     -     7.00   138.19   140.54     521  ! BG458009 NF037B11F1_PF1096 Phos
qb_est102:BG591958     -     7.00   138.18   140.76     522  ! BG591958 EST499800 P. intestans
qb_est53:AW906526            7.00   138.04   143.23     533  ! AW906526 EST542648 potato stolo
qb_est81:BE920425      -     7.00   137.97   144.67     539  ! BE920425 EST424194 potato leave
qb_est87:BF434054      -     7.00   137.95   144.99     540  ! BF434054 7o98b09.x1 NCI_CGAP_Co
qb_qss33:AZ837984      +     7.00   137.83   147.19     550  ! AZ837984 2M0134D03R Mouse 10kb
qb_est53:AW964661      +     7.00   137.43   154.33     584  ! AW964661 EST476744 MAGE resequ
qb_est74:BE441916      -     7.00   137.21   159.45     604  ! BE441916 925009E04.x1 C. reinha
qb_est81:BE924525      -     7.00   137.14   160.80     610  ! BE924525 EST428210 potato leave
qb_qss5:FR0004867      +     7.00   137.04   162.82     619  ! Z88677 F.rubripes GSS sequence,
qb_est99:BG450529      -     7.00   136.49   177.18     683  ! BG450529 093e05 Mature tuber la
qb_est99:BG350514      -     7.00   136.21   181.30     702  ! BG350514 094A01 Mature tuber la
qb_est41:AV753114      -     7.00   136.01   185.92     723  ! AV753114 AV753114 NPD Homo sapi
qb_est102:BG590968     -     7.00   135.80   190.96     746  ! BG590968 EST498810 P. intestans
qb_est97:BG179568      -     7.00   135.46   202.05     797  ! BG179568 602428182F1 NIH_MGC_91
qb_est53:AW906709      -     7.00   135.42   203.14     802  ! AW906709 EST542834 potato st
qb_qss8:CNS03J8C       -     7.00   135.07   209.61     832  ! AL222984 Tetraodon nigroviri
qb_est30:AU124989      +     7.00   134.90   214.33     854  ! AU124989 AU124989 NT2RM4 Hom
qb_qss4:CNS04J3G       -     7.00   134.85   215.83     861  ! AL424475 Tetraodon nigroviri
qb_est76:BE560990      +     7.00   134.84   216.05     862  ! BE560990 601437227F1 NIH_MGC


seq_name: qb_est80:BE871340

 seq_fasta_offset: 0
LOCUS       BE871340         945 bp    mRNA            EST       20-SEP-2000
DEFINITION  601449457F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3853334 5',
            mRNA sequence.
ACCESSION   BE871340
VERSION     BE871340.1  GI:10320116
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 945)
  AUTHORS   NIH_MGC_Project.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cmaps@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9577  row: a  column: 15
            High quality sequence stop: 656.
FEATURES             Location/Qualifiers
     source          1..945
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3853334"
                     /clone_lib="NIH_MGC_65"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer:  Xho dT.
                     Average insert size 1.8 kb.  Library constructed by Life
                     Technologies. "
BASE COUNT       170 a     363 c     300 g     112 t
ORIGIN


alignment_scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_84_92 x BE871340  ..

 Align seq 1/1  to: BE871340  from: 1  to: 945

        1 GlnGlyArgProArgGlnProGlyGln 9
          |||||||||||||||||||||||||||
      609 CAGGAGCCACCTGAGGAGCCTGAGCAG 635

seq_name: qb_est97:BG179568

seq_documentation_block:
LOCUS       BG179568         991 bp    mRNA            EST       06-FEB-2001
DEFINITION  602428182F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4429913 5',
            mRNA sequence.
ACCESSION   BG179568
VERSION     BG179568.1  GI:12686271
```

```
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 991)
  AUTHORS   NIH_MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at
            http://image.llnl.gov
            Plate: LLAM10181  row: d  column: 22
            High quality sequence stop: 655.
FEATURES             Location/Qualifiers
     source          1..991
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1470415"
                     /clone_lib="NIH_MGC_91"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage resistant)"
                     /note="Organ: prostate; Vector: pCMV_SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo dT primed.
                     Average insert size 1.4 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC library."
BASE COUNT      190 a    351 c    305 g    145 t
ORIGIN


alignment scores:
            Quality:     9.00            Length:       9
              Ratio:    1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2 copy_84_92 x BG179568   ..

 Align seq 1/1  to: BG179568  from: 1  to: 991

        1 GluGluProProGluGluProGluGln 9
          ||||||||||||||||||||||||||||
      154 GAGGAGCCACCTGAGGAGCCTGAGCAG 179

seq name: qb|qss28:AZ511504

seq documentation block:
LOCUS       AZ511504       509 bp    DNA             GSS       05-OCT-2000
DEFINITION  IM0356N19F Mouse 10kb plasmid UHGCIM library Mus musculus genomic
            clone UHGCIM0356N19 F, DNA sequence.
ACCESSION   AZ511504
VERSION     AZ511504.1  GI:10692736
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 509)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beaconn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,L., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., Slc, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert length: 10000  Std Error: 0.00
            Plate: 0356  row: N  column: 19
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 509.
FEATURES             Location/Qualifiers
     source          1..509
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UHGCIM0356N19"
                     /clone_lib="Mouse 10kb plasmid UHGCIM library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10 Gold, Tl resistant, F'"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained from the Jackson
                     Laboratory Mouse DNA Resource
                     [illegible] the DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high molar excess. The
                     adaptored DNA was purified and size selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      113 a    118 c    130 g    148 t
ORIGIN


alignment scores:
            Quality:     8.00            Length:       8
              Ratio:    1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2 copy_84_92 x AZ511504   ..

 Align seq 1/1  to: AZ511504  from: 1  to: 509

        2 GluProProGluGluProGluGln 9
          ||||||||||||||||||||||||||
      335 GAGCCACCTGAGGAGCCTGAGCAG 359

seq name: qb|qss26:AZ390410

seq documentation block:
LOCUS       AZ390410       560 bp    DNA             GSS       03-OCT-2000
DEFINITION  IM0151E16F Mouse 10kb plasmid UHGCIM library Mus musculus genomic
            clone UHGCIM0151E16 F, DNA sequence.
ACCESSION   AZ390410
VERSION     AZ390410.1  GI:10505453
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 560)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beaconn,T., Duval,B., Hamil,C.,
```

```
            Islam,H., Longacre,S., Mahmoud,M., Meemeh,E., Pedersen,I., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tinsley,A., Von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000  Std Error: 0.00
            Plate: 0151  row: E  column: 16
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 560.
FEATURES             Location/Qualifiers
     source          1..560
                     /organism "Mus musculus"
                     /strain "C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone "UUGCIM0151E16"
                     /clone_lib "Mouse [illegible] UUGCIM Library"
                     /sex="Male"
                     /lab_host "E. Coli strain XL10-Gold, Tet resistant, F-"
                     /note "Vector: PWD42nx; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/imr/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4742114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10 Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       159 a    172 c    125 g    104 t
ORIGIN


alignment_scores:
             Quality:     8.00              Length:       8
               Ratio:    1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_84_92 x AA990410/rev  ..

 Align seq 1/1  to reverse of: AA990410  from: 1  to: 560

        2 GluProProGluGluProGluGln 9
          |||||||||||||||||||||||||
      511 GAGCCGCCCGAGGAGCCGGAACAG 488

seq_name: qb_est80:BE844192

seq_documentation_block:
LOCUS        BE844192       577 bp    mRNA             EST       22-SEP-2000
DEFINITION   RC3-TN0094-110800-021-g08 TN0094 Homo sapiens cDNA, mRNA sequence.
ACCESSION    BE844192
VERSION      BE844192.1  GI:10276570
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 577)
  AUTHORS    Dias Neto,E., Garcia Correa,R., Verjovski Almeida,S., Briones,M.R.,
             Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
             Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
             Brumbstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., Hiro
             ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
             Simpson,A.J.
  TITLE      Shotgun sequencing of the human transcriptome with ORF expressed
             sequence tags
  JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE    20202663
COMMENT      Contact: Simpson A.J.G.
             Laboratory of Cancer Genetics
             Ludwig Institute for Cancer Research
             Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo SP,
             Brazil
             Tel: 55-11-2784922
             Fax: 55 11 2767001
             Email: aclimpson@ludwig.org.br
             This sequence was derived from the LICR/ONCO Human Cancer Genome
             Project. This entry can be seen in the following URL
             http://www.ludwig.org.br/ESTs/pet?tm2 plus TN0094 110
             800-021 g08&t8 2000-08-11&t4-1)
             Seq primer: puc 18 forward
             High quality sequence start: 20
             High quality sequence stop: 572.
FEATURES             Location/Qualifiers
     source          1..577
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone_lib "TN0094"
                     /dev_stage "Adult"
                     /note "organ: testis_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  CRESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       129 a    149 c    146 g    151 t      2 others
ORIGIN


alignment_scores:
             Quality:     8.00              Length:       8
               Ratio:    1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_84_92 x BE844192_rev  ..

 Align seq 1/1  to reverse of: BE844192  from: 1  to: 577

        2 GluProProGluGluProGluGln 9
          |||||||||||||||||||||||||
      564 GAGCCGCCCGAGGAGCCGGAACAG 541

seq_name: qb_est80:BE844191

seq_documentation_block:
LOCUS        BE844191       583 bp    mRNA             EST       22-SEP-2000
DEFINITION   RC3-TN0094-110800-021-g02 TN0094 Homo sapiens cDNA, mRNA sequence.
ACCESSION    BE844191
VERSION      BE844191.1  GI:10276569
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 583)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brumstein,A., de Oliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202668
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo SP,
            Brazil
            Tel: +55 11 2704922
            Fax: +55 11 2782801
            Email: asimpson@ludwig.org.br
            This Sequence was derived from the FANTOM/RIKEN Mouse cDNA
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/cgi-bin/ecbot.pl?fl=mr,s=7292&el=1-
            800-021 q02&t3 2000-08-11&t4 1)
            Seq primer: puc18 forward
            High quality sequence start: 28
            High quality sequence stop: 583.
FEATURES             Location/Qualifiers
     source          1..583
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="TN0094"
                     /dev_stage="Adult"
                     /note="Organ: testis_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini library was made by cloning products
                     derived from ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 - Ludwig Institute for Cancer Research)
                     profiles into the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      141 a    151 c    146 g    154 t      1 others
ORIGIN


alignment scores:
            Quality:     8.00            Length:       8
              Ratio:    1.000              Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2 (COPY 84_92 x BE844191/rev  ..

 Align seq 1/1  to reverse of: BE844191  from: 1  to: 583

       1 GlnGlnProProGluGluProGlu 8
         ||||||||||||||||||||||||
     244 GAAGAGCCACCAGAAGAGCCGGAG 221

Seq name: qb qss9:AQ272338

Seq documentation block:
LOCUS       AQ272338       588 bp    DNA             GSS       03-NOV-1998
DEFINITION  nbxb0027H18r CUGI Rice BAC Library Oryza sativa genomic clone
            nbxb0027H18r, DNA sequence.
ACCESSION   AQ272338
VERSION     AQ272338.1  GI:4825654
KEYWORDS    GSS.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 588)
  AUTHORS   Wing,R.A. and Dean,R.A.
  TITLE     A BAC End Sequencing Framework to Sequence the Rice Genome
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: CGAAACAGCTATGACCATG
            Class: BAC ends
            High quality sequence start: 7
            High quality sequence stop: 597.
FEATURES             Location/Qualifiers
     source          1..588
                     /organism="Oryza sativa"
                     /cultivar="Japonica"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="nbxb0027H18r"
                     /clone_lib="CUGI Rice BAC Library"
                     /dev_stage="leaf"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII; Rice is one of two most popular grains in the
                     world. Half of the world population especially those
                     inhabiting highly populated areas of the humid tropics
                     and subtropics, rely on rice as their primary source of
                     carbohydrate. Monocotyledonous rice is a diploid plant
                     (2n=24) with a haploid genome equivalent of 431 Mbp
                     (Arumuganathan and Earle, 1991). The relatively small
                     genome of rice, three times larger than that of
                     Arabidopsis, makes it suitable for genomic studies. In
                     order to facilitate positional cloning, physical mapping
                     and genome sequencing of rice, we have constructed a BAC
                     library from Oryza sativa, Nipponbare variety. The
                     library contains 86,864 clones with an average insert size
                     of 128.5 Kb providing 10.9 haploid genome equivalents. The
                     deep coverage allows the isolation a particular sequence
                     with a probability of 99.9 %. Two high density filters,
                     each containing 18,432 clones (doubly spotted), represent
                     the whole library for colony screening."
BASE COUNT      135 a    202 c    167 g     79 t      5 others
ORIGIN


alignment scores:
            Quality:     8.00            Length:       8
              Ratio:    1.000              Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2 (COPY 84_92 x AQ272338  ..

 Align seq 1/1  to: AQ272338  from: 1  to: 588

       1 GlnGlnProProGluGluProGlu 8
         ||||||||||||||||||||||||
     519 GAAGAGCCACCCGAGGAACCGGAA 542

Seq name: qb est102:BG598416

Seq documentation block:
LOCUS       BG598416       122 bp    mRNA            EST       12-APR-2001
DEFINITION  EST353316 cDNA3 Solanum tuberosum cDNA clone cDNA3ESZb23 5', mRNA
            sequence.
ACCESSION   BG598416
VERSION     BG598416.1  GI:13618257
KEYWORDS    EST.
SOURCE      potato.
  ORGANISM  Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
```

```
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum
REFERENCE   1  (bases 1 to 122)
  AUTHORS   van der Hoeven,R., Bezzerides,J., Sun,H., Chu,J., Chiominto,A.,
            Bougri,O., Buell,C.R., Ronning,C., Tanksley,S. and Baker,B.
  TITLE     Generations of ESTs from sprouting potato eyes
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cathy Ronning
            The Institute for Genomic Research
            For clone info: please contact Research Genetics, Libraries
            Division tel 1-800-711-6195, email cdna@resgen.com
            Seq primer: M13F-R.
FEATURES             Location/Qualifiers
     source          1..122
                     /organism="Solanum tuberosum"
                     /cultivar="Kennebec"
                     /db_xref="taxon:4113"
                     /clone="cSTS21023"
                     /clone_lib "cSTS"
                     /tissue_type "sprouting eyes from tubers"
                     /dev_stage "12-14 weeks post harvest"
                     /lab_host="SOLR"
                     /note "Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
                     XhoI; Various sizes of sprouting eyes (2mm to 15mm) were
                     taken from tubers. The tubers were incubated at 20C in the
                     dark for 2-3 weeks prior to sprouting. The eyes were
                     frozen in liquid nitrogen immediately upon removal from
                     tubers."
BASE COUNT        30 a      22 c      33 g      37 t
ORIGIN


alignment_scores:
             Quality:     7.00              Length:       7
               Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-9458-2_COPY_84_92 x BG598416/rev    ..

 Align Seq 1/1  to reverse of: BG598416  from: 1  to: 122

        1 GluGluProProGluGluPro 7
          |||||||||||||||||||||
       59 GAAGAGCCPCCAGAAGAGCCT 39

seq_name: gb_est70:BE083507

seq_documentation_block:
LOCUS       BE083507       312 bp    mRNA            EST       12-JUN-2000
DEFINITION  RC5 BT0645-110400-013-A11 BT0645 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE083507
VERSION     BE083507.1  GI:8473846
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 312)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., de Oliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55 11-2704922
            Fax: +55 11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            http://www.ludwig.org.br/cgi-bin/db_hits.pl?id=BT0645-110
            400-013-A11&lib=2000-04-11&lt4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 311.
FEATURES             Location/Qualifiers
     source          1..312
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib "bt0645"
                     /dev_stage "Adult"
                     /note "Organ: breast, Vector: puc18, Site 1: SmaI; Site 2:
                     SmaI; A mini-library was made by cloning products derived
                     from  cDNESIS PCR (U.S. Letters Patent application No. 196
                     ,716 - Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT        48 a     107 c      84 g      73 t
ORIGIN


alignment_scores:
             Quality:     7.00              Length:       7
               Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-9458-2_COPY_84_92 x BE083507/rev    ..

 Align Seq 1/1  to reverse of: BE083507  from: 1  to: 312

        1 GluGluProProGluGluPro 7
          |||||||||||||||||||||
       54 GAAGAGCCPCCAGAAGAGCCT 34

seq_name: gb_est78:BE707851

seq_documentation_block:
LOCUS       BE707851       333 bp    mRNA            EST       12-SEP-2000
DEFINITION  QV4-HT0544-29Aug00-268-q03 HT0544 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE707851
VERSION     BE707851.1  GI:10096116
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 333)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., de Oliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55 11-2704922
            Fax: +55 11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
```

```
                (http://www.ludwig.org.br/orestes/...pl?...
                700 268 q03&t3 2000 07 29&t4 1)
                Seq primer: puc 18 forward
                High quality sequence start: 26
                High quality sequence stop: 333.
FEATURES             Location/Qualifiers
     source          1..333
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone_lib "HT0544"
                     /dev_stage "Adult"
                     /note "Organ: head neck; Vector: puc18; Site 1: SmaI;
                     Site_2: SmaI; A mini library was made by cloning products
                     derived from ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 - Ludwig Institute for Cancer Research)
                     profiles into the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       68 a     63 c     98 g    104 t
ORIGIN


alignment scores:
           Quality:     7.00          Length:       7
             Ratio:    1.000            Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_84_92 x BE707851/rc ...

 Align seq 1/1  to reverse of: BE707851  from: 1  to: 333

        1 GluGluProProGluGluPro 7
          |||||||||||||||||||||
       39 GAAGAGCCCCCGAAGAGCCA 19

seq_name: gb_est110:W63245

seq_documentation_block:
LOCUS       W63245         370 bp    mRNA            EST       22-MAY-2000
DEFINITION  TgESTzy68h01.r1 TgME49 Tachyzoite cDNA Toxoplasma gondii cDNA clone
            tgzy68h01.r1 5' similar to SW:RL1A_XENLA P08429 60S RIBOSOMAL
            PROTEIN L1A. [1] ;, mRNA sequence.
ACCESSION   W63245
VERSION     W63245.1  GI:1370885
KEYWORDS    EST.
SOURCE      Toxoplasma gondii.
  ORGANISM  Toxoplasma gondii
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
            Sarcocystidae; Toxoplasma.
REFERENCE   1  (bases 1 to 370)
  AUTHORS   Hehl,A., Manger,I., Marra,M., Sibley,L.D., Ajioka,J.A., Aslett,M.A.
            , Dietrich,N., Dubuque,T., Hillier,L., Kucaba,T., Wan,K.L.,
            Waterston,R.H. and Boothroyd,J.
  TITLE     WashU Merck Stanford NIH Toxoplasma EST project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: toxo@watson.wustl.edu
            David Sibley at toxoest@borcim.wustl.edu for further information
            relating to organism, clone or library availability.
            Seq primer: T3
            High quality sequence stop: 294.
FEATURES             Location/Qualifiers
     source          1..370
                     /organism "Toxoplasma gondii"
                     /strain "ME49, clone PDS"
                     /db_xref "taxon:5811"
                     /clone "tgzy68h01.r1"
                     /clone_lib "TgME49 Tachyzoite cDNA"
                     /lab_host "XL1-Blue MRF'"
                     /note "Vector: Lambda ZAP; Site 1: EcoRI; Site 2: XhoI;
                     Toxoplasma parasites were grown in human foreskin
                     fibroblasts cultures in vitro. The library was
                     constructed by A. Hehl and I. Manger, Stanford University.
                     cDNAs were synthesized from polyA mRNAs by oligo d(T)
                     priming and directionally cloned into the EcoRI and XhoI
                     sites of the Lambda Zap vector using the ZAP cDNA
                     synthesis Kit (Stratagene). Warning: the library contains a
                     small percentage of human cDNAs derived from the human
                     host cells."
BASE COUNT       72 a    119 c    107 g     58 t     14 others
ORIGIN


alignment scores:
           Quality:     7.00          Length:       7
             Ratio:    1.000            Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_84_92 x W63245 ...

 Align seq 1/1  to: W63245  from: 1  to: 370

        1 GluGluProProGluGluPro 7
          |||||||||||||||||||||
      227 GAAGAACCTCCTGAAGAACCA 247

seq_name: gb_est106:N60773

seq_documentation_block:
LOCUS       N60773         395 bp    mRNA            EST       22-MAY-2000
DEFINITION  TgESTzy21d08.r1 TgRH Tachyzoite cDNA Toxoplasma gondii cDNA clone
            tgzy21d08.r1 5', mRNA sequence.
ACCESSION   N60773
VERSION     N60773.1  GI:1206924
KEYWORDS    EST.
SOURCE      Toxoplasma gondii.
  ORGANISM  Toxoplasma gondii
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
            Sarcocystidae; Toxoplasma.
REFERENCE   1  (bases 1 to 395)
  AUTHORS   Hehl,A., Manger,I., Marra,M., Sibley,L.D., Ajioka,J.A., Aslett,M.A.
            , Dietrich,N., Dubuque,T., Hillier,L., Kucaba,T., Wan,K.L.,
            Waterston,R.H. and Boothroyd,J.
  TITLE     WashU Merck Stanford NIH Toxoplasma EST project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: toxo@watson.wustl.edu
            David Sibley at toxoest@borcim.wustl.edu for further information
            relating to organism, clone or library availability.
            Seq primer: T3
            High quality sequence stop: 284.
FEATURES             Location/Qualifiers
     source          1..395
                     /organism "Toxoplasma gondii"
                     /strain "RH"
                     /db_xref "taxon:5811"
                     /clone "tgzy21d08.r1"
                     /clone_lib "TgRH Tachyzoite cDNA"
                     /lab_host "XL1-Blue MRF'"
                     /note "Vector: Lambda ZAP; Site 1: EcoRI; Site 2: XhoI;
                     Toxoplasma RH strain tachyzoites were grown in human
                     foreskin fibroblast cultures in vitro. The library was
                     constructed by K.L. Wan, Cambridge University. cDNAs were
                     synthesized from polyA RNAs by oligo d(T) priming and
```

directionally cloned into the EcoRI to XhoI sites of the
Lambda ZapII vector using the ZAP-cDNA synthesis kit
(Stratagene). WARNING: the library contains a small
percentage of cDNAs derived from the human host cells."
BASE COUNT 81 a 124 c 103 g 63 t 24 others
ORIGIN

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_84_92 x N60773 ..

Align seq 1/1 to: N60773 from: 1 to: 395

1 GluGluProProGluGluPro 7
|||||||||||||||||||||
242 GAAGAACCTCCTGAAGAACCA 262

Seq_name: gb_est102:BG554981

Seq_documentation_block:
LOCUS BG554981 397 bp mRNA EST 09-APR-2001
DEFINITION dac34108.y1 NICHD_XGC_Hel Xenopus laevis cDNA clone IMAGE:4408863
5' similar to TR:Q15022 Q15022 ORF 1, mRNA sequence.
ACCESSION BG554981
VERSION BG554981.1 GI:18566761
KEYWORDS EST
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 397)
AUTHORS Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Billier,L., Pape,D.
, Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,K.,
Waterston,R. and Wilson,R.
TITLE WashU Xenopus EST project, 1999
JOURNAL Unpublished (1999)
COMMENT Contact: Sandy Clifton, Ph.D.
WashU Xenopus EST project Tel:
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, Mo 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library constructed by Life Technologies. DNA Sequencing by:
Washington University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
Seq primer: -40RP from Gibco.
FEATURES Location/Qualifiers
source 1. .397
/organism "Xenopus laevis"
/db_xref "taxon:8355"
/clone="IMAGE:4408863"
/clone_lib "NICHD_XGC_Hel"
/lab_host "DH10B (phage-resistant)"
/note="Organ: heart; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally, Primer: Oligo dT.
Average insert size 1.6 kb. Constructed by Life
Technologies."
BASE COUNT 135 a 90 c 95 g 77 t
ORIGIN

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_84_92 x BG554981 ..

Align seq 1/1 to: BG554981 from: 1 to: 397

1 GluGluProProGluGluPro 7
|||||||||||||||||||||
94 GAAGAACCTCCTGAAGAACCA 114

Seq_name: gb_est93:BF868534

Seq_documentation_block:
LOCUS BF868534 412 bp mRNA EST 17-JAN-2001
DEFINITION IL8_EI0115-121000-284-C06 EI0115 Homo sapiens cDNA, mRNA sequence
ACCESSION BF868534
VERSION BF868534.1 GI:12258664
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 412)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brumbley,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55 11 2707005
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.licr.org.br/scripts/perlcrm?2 pl?11 IL8&12 IL--EI0115
121000-284-C06&13 2000-10-12&14 1)
Seq primer: puc 18 forward
High quality sequence stop: 393.
FEATURES Location/Qualifiers
source 1. .412
/organism "Homo sapiens"
/db_xref "taxon:9606"
/clone_lib "EI0115"
/dev_stage "Adult"
/note "Organ: lung_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,710 Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 88 a 142 c 135 g 65 t
ORIGIN

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_84_92 x BF868534 ..

```
 Align seq 1/1  to: BE868534  from: 1  to: 412

        3 ProProGluGluProGluGln 9
          ||||||||||||||||||||
      282 CCCCCCGAGGAGCCTGAGCAG 302

seq name: qb ost97:BG165837

seq documentation block:
LOCUS       BG165837        417 bp    mRNA            EST       06 FEB 2001
DEFINITION  602344435F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4454485 5',
            mRNA sequence.
ACCESSION   BG165837
VERSION     BG165837.1  GI:12672553
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 417)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10246  row: e  column: 10
            High quality sequence stop: 329.
FEATURES             Location/Qualifiers
     source          1..417
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4454485"
                     /clone_lib="NIH_MGC_89"
                     /tissue_type="hypernephroma, cell line"
                     /lab_host="DH10B (phage resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site 1: NotI;
                     Site 2: SalI; Cloned unidirectionally; oligo dT primed.
                     Average insert size 1.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT        78 a    145 c    143 g     61 t
ORIGIN


alignment scores:
            Quality:     7.00            Length:       7
              Ratio:    1.000              Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2-COPY_84_92 x BG165837  ..

 Align seq 1/1  to: BG165837  from: 1  to: 417

        1 GlnGluProProGluGluPro 7
          |||||||||||||||||||
      254 GAGGAGCCCACCCCAGGAGCCG 274
```

```
OM of: US-09-126-945B-2_COPY_130_138 to: GenEmbl:*  out_format : pts

Date: Sep 14, 2001  9:37 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV xlp
 -Q "/cgn2_6.3sp4/qy_s-pto/s/sq1_2/pto/tmp/tmp_11%2881_14%211_10%3/qy_1.pto; fasta_1 %3F
 -DB GenEmbl -QFMT=fastap -SUFFIX olip2n.rge  GAPOP 4.500
 -GAPEXT 0.050 -MINMATCH=0.100 -LOOPCL 0.000  LOOPEXT 0.000
 -QGAPOP 4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT 60.000
 -YGAPOP=6.000 -YGAPEXT=7.000 -YGAPOP=60.000  YGAPEXT 60.000
 -DELOP=6.000 -DELEXT=7.000 -START 1 -MATRIX oligo
 -TRANS human40.cdi -LIST=45 -DOCALIGN=200  THR_SCORE quality
 -THR_MIN=1 -ALIGN 15 -MODE=LOCAL -OUTFMT pts  NORM OAT  MINLEN 0
 -MAXLEN=2000000000 -USER=US09126945_ACCN1_1_7594 -NCPU 6
 -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS 1

Search information block:
Query: US-09-126-945B-2_COPY_130_138
Query length: 9
Database: GenEmbl:*
Database sequences: 1344157
Database length: -856060004
Search time (sec): 6503.700000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000


score_list:
Sequence                 Strd Orig    ZScore    EScore Len  ! Documentation ..
qb_rol:AB019436            +     9.00  179.13   0.0968    1704  ! AB019436 Mus musculus mRNA for
qb_pat1:AX048148           +     9.00  178.42   0.1061    1894  ! AX048148 Sequence 1 from Paten
qb_pr4:AF071538            +     9.00  178.42   0.1061    1894  ! AF071538 Homo Sapiens Ets tran
qb_pr1:AB031549            +     9.00  178.40   0.1063    1898  ! AB031549 Homo sapiens PSP mRNA
qb_pat1:AX048159           +     9.00  174.64   0.1729    2317  ! AX048159 Sequence 12 from Pate
qb_pr6:AL157372            -     9.00  148.94   4.65      15660  ! AL157372 Human DNA sequence
qb_pat2:109074             +     7.00  142.31  10.89      885  ! 109074 Sequence 1 from Patent W
qb_pl3:NEUMUS8             -     7.00  136.75  22.19      2419  ! U28472 Neurospora crassa DNA
qb_pl2:AN1291452           -     7.00  130.64  48.59      5030  ! AJ291452 Aspergillus nidulans
qb_vi1:AB042808            +     7.00  127.74  70.53      7697  ! AB042808 Chiba virus genomic R
qb_vi2:S57540              +     7.00  126.90  78.57      8717  ! S57540 orf RF1...orf UL2 [huma
qb_vi2:HHV6IMMEAPB         +     7.00  116.90  283.22     30492  ! L25528 Human herpesvirus 6 U
qb_htq18:AC087855          +     7.00  113.69  427.46     61958  ! AC087855 Homo sapiens chromos
qb_htq2:AC010861           +     7.00  113.54  435.75     63369  ! AC010861 Homo sapiens clone R
qb_htq5:AC016176           -     7.00  112.38  505.51     75254  ! AC016176 Homo sapiens clone R
qb_htq1:AC008611           -     7.00  109.32  748.43     118528  ! AC008611 Homo sapiens chromo
qb_pr9:HSDJ756N5           -     7.00  109.03  777.00     123782  ! AL132825 Human DNA sequence
qb_htq1:AC008684           -     7.00  108.14  870.24     141132  ! AC008684 Homo sapiens 5,
qb_htq3:AC011792           -     7.00  107.71  919.80     150467  ! AC011792 Homo sapiens clone
qb_htq14:AC068619          -     7.00  107.60  933.56     153089  ! AC068619 Homo sapiens chromo
qb_vi2:HHV6AGNM            +     7.00  107.33  966.30     159321  ! X83413 Human herpesvirus-6 (
qb_vi1:AB021506            +     7.00  107.23  978.09     161573  ! AB021506 Human herpesvirus 6
qb_htq10:AC025546          +     7.00  107.22  979.94     161927  ! AC025546 Homo sapiens chromo
qb_vi1:AF157706            +     7.00  107.21  980.92     162114  ! AF157706 human herpesvirus
qb_pr2:AC006083            +     7.00  107.19  983.59     162625  ! AC006083 Homo sapiens chroma
qb_pr4:AC025613            -     7.00  106.79  1.0e+03    172475  ! AC025613 Homo sapiens chromo
qb_htq24:AP003484          -     7.00  106.56  1.1e+03    178665  ! AP003484 Homo sapiens chromo
qb_htq23:AP000853          +     7.00  106.54  1.1e+03    179137  ! AP000853 Homo sapiens chroma
qb_htq18:AC084355          +     7.00  106.43  1.1e+03    182098  ! AC084355 Homo sapiens chromo
qb_htq14:AC068054          -     7.00  106.35  1.1e+03    184284  ! AC068054 Homo sapiens chromo
qb_pr3:AC010583            -     7.00  106.34  1.1e+03    184569  ! AC010583 Homo sapiens chromo
qb_htq16:AC079123          -     7.00  106.25  1.1e+03    187087  ! AC079123 Homo sapiens chromo
qb_htq24:AP003327          +     7.00  106.23  1.1e+03    187499  ! AP003327 Homo sapiens chromo
qb_pl2:AP000836            +     7.00  106.14  1.1e+03    190014  ! AP000836 oryza sativa genomi
qb_htq21:AL456489          +     7.00  105.97  1.2e+03    195048  ! AL456489 Homo sapiens chromo
qb_pr5:AF165124            +     7.00  105.94  1.2e+03    195909  ! AF165124 Homo sapiens chroma
qb_htq19:AF286884          +     7.00  105.60  1.2e+03    205900  ! AF286884 Homo sapiens chroma
qb_htq18:AC087671          +     7.00  105.59  1.2e+03    206246  ! AC087671 Homo sapiens chromo
qb_pr10:HTCRBCHR9          +     7.00  105.27  1.3e+03    216293  ! AF029998 Homo sapiens chroma
qb_htq13:AC067717          +     7.00  105.02  1.3e+03    224319  ! AC067717 Homo sapiens chr
qb_pr1:AC002523            +     7.00  104.90  1.3e+03    228516  ! AC002523 Homo sapiens Xq2
qb_htq21:AL450573          -     7.00  104.84  1.5e+03    267581  ! AL450573 Homo sapiens chr
qb_sts2:G29979             -     6.00  129.33  65.24      380  ! G29979 EST64 tomato root
qb_pat1:AX014497           -     6.00  127.51  72.63      373  ! AX014497 Sequence 53 from Pa
qb_pat2:E64921             -     6.00  127.51  72.63      373  ! E64921 cDNA sequence of gene

seq_name: qb_rol:AB019436

 s [Genome:Gaff = 1]  k:
LOCUS       AB019436      1704 bp    mRNA            ROD       20-JAN-2000
DEFINITION  Mus musculus mRNA for prostate specific transcription factor ets,
            complete cds.
ACCESSION   AB019436
VERSION     AB019436.1  GI:4760536
KEYWORDS    Prostate specific transcription factor ets; mPse.
SOURCE      Mus musculus (strain:C57BL/6) Adult male Colon Epithelial cDNA to
            mRNA.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Yamada,N., Tamai,Y., Miyamoto,H. and Nozaki,M.
  TITLE     Cloning and expression of the mouse Pse gene encoding a novel Ets
            family member
  JOURNAL   Gene 241 (2), 267-274 (2000)
  MEDLINE   20137504
REFERENCE   2  (bases 1 to 1704)
  AUTHORS   Nozaki,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-NOV-1998) to the EMBL/GenBank/DDBJ databases. Masumi
            Nozaki, Osaka University, Institute for Microbial Diseases; 3-1,
            Yamadaoka, Suita, Osaka 565-0871, Japan
            (E-mail:mnozaki-biken.osaka-u.ac.jp, Tel:81-6-879-8338,
            Fax:81-6-879-8339)
FEATURES             Location/Qualifiers
     source          1..1704
                     /organism="Mus musculus"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /cell_type="Epithelial"
                     /dev_stage="Adult"
                     /sex="male"
                     /tissue_type="Colon"
     gene            362..1339
                     /gene="mPse"
     CDS             362..1339
                     /gene="mPse"
                     /codon_start=1
                     /product="Prostate specific transcription factor ets"
                     /protein_id="BAA77329.1"
                     /db_xref="GI:4760537"
                     /translation="MCSASPCFSNVSFSQLIEPPVAPQLTPKAASCAMGPEKQPWS
                     PSPPATPEQCLSAFYLSYFNMYPPDSSWVAKVPEAKAGFDHPEFEEEQCPVIDSQASGS
                     TLDEBSIEQVQSMVVGEVLKDTEIACKTLNTIADSGDWSFGNVQKWLIWIEHGYEIPP
                     ACKAFQETGGKPLCAMSEFQPRQRAPLCGDVIHAHLDIWKSAAWMKEDISFGTLHYCA
                     STSEEGWIDGEVDSSCSGQPTHLWQFLKELLIKPHSYGPETPWFLNKFCIEKTFSSAQ
                     VAELWGVEENPFAMAYDELSESTPQAYYEPCLIEPPDISQKLVYQFVHPV"
     polyA_site      1704
                     /note="10 a nucleotides"
BASE COUNT       369 a    527 c    491 g    317 t
ORIGIN


alignment_scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_130_138 x AB019436  ..

 Align_seq 1/1  to: AB019436  from: 1  to: 1704
```

```
        1 ValLeuLysAspIleGluThrAlaCys 9
          ||||||||||||||||||||||||||||
      719 GTCCTGAAAGATATTGAGACGGCCTGC 745

seq name: qb pat1:AX048148

seq documentation block:
LOCUS       AX048148      1894 bp    DNA             PAT       15-DEC-2000
DEFINITION  Sequence 1 from Patent WO00/00092.
ACCESSION   AX048148
VERSION     AX048148.1  GI:11876967
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo
REFERENCE   1  (bases 1 to 1894)
  AUTHORS   Kaufmann,J. and Xin,H.
  TITLE     Expression of ets domain proteins in cancer
  JOURNAL   Patent: WO 0070092 A 1 23-NOV 2000;
            Chiron Corporation (US)
FEATURES             Location/Qualifiers
     source          1..1894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             416..1423
                     /note="human JKETS."
                     /codon_start=1
                     /protein_id="AAL99092.1"
                     /db_xref="GI:11876968"
                     /translation="MGSASPGLSSVSPSHLLEPPDIVSKTGLEKAAAGAVGLERMDWS
                     PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAPGASSKEEPPEEFEQCPVTDSQAPA
                     GSLDLVPGGLTLEEHSLEQVQSMVVGEVILKDTETACKLLNITADPMDWSPSNVQKWLL
                     WTEHQYRLPPMGKAFQELAGKELCAMSEEQFKQKSPLGGDVLHAHLDTWKSAAWMKER
                     TSPGATHYCASISEESWTDSEVDSSCSGQPTHLWQFLKELLLKPHSYGRFIRWLNEEK
                     GIFKIEDSAQVARLWGIRKNKPAMNYDKLSRSIRQYYEKGIIEKPDISQPLVYQFVHP
                     I"
BASE COUNT       468 a    654 c    571 g    302 t
ORIGIN


alignment scores:
            Quality:     9.00              Length:       9
              Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 DS-09-126-945b-2_copy_130_138 x AX048148   ..

 Align seq 1/1  to: AX048148  from: 1  to: 1894

        1 ValLeuLysAspIleGluThrAlaCys 9
          ||||||||||||||||||||||||||||
      804 GTGCTCAAGGACATCGAGACGGCCTGC 829

seq name: qb pr4:AF071538

seq documentation block:
LOCUS       AF071538      1894 bp    mRNA            PRI       10-JAN-2000
DEFINITION  Homo sapiens Ets transcription factor FDEF (PDEF) mRNA, complete
            cds.
ACCESSION   AF071538
VERSION     AF071538.1  GI:4007417
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1894)
  AUTHORS   Oettgen,P., Finger,E., Sun,Z., Akbarali,Y., Thamrongsak,U.,
            Boltax,J., Grall,F., Dube,A., Weiss,A., Brown,L., Quinn,G., Kas,K.,
            Endress,G., Kunsch,C. and Libermann,T.A.
  TITLE     PDEF, a novel prostate epithelium-specific ets transcription
            factor, interacts with the androgen receptor and activates
            prostate specific antigen gene expression
  JOURNAL   J. Biol. Chem. 275 (2), 1216-1225 (2000)
  MEDLINE   20092894
REFERENCE   2  (bases 1 to 1894)
  AUTHORS   Libermann,T.A., Oettgen,P., Kunsch,C. and Finger,E
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUN-1998) Medicine, Beth Israel Deaconess Medical
            Center, 330 Brookline Ave., Boston, MA 02215, USA
FEATURES             Location/Qualifiers
     source          1..1894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6p21.3"
                     /tissue_type="prostate"
     gene            1..1894
                     /gene="PDEF"
     CDS             416..1423
                     /gene="PDEF"
                     /note="prostate epithelium specific"
                     /codon_start=1
                     /product="Ets transcription factor PDEF"
                     /protein_id="AAC95296.1"
                     /db_xref="GI:4007418"
                     /translation="MGSASPGLSSVSPSHLLEPPDIVSKTGLEKAAAGAVGLERMDWS
                     PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAPGASSKEEPPEEFEQCPVTDSQAPA
                     GSLDLVPGGLTLEEHSLEQVQSMVVGEVILKDTETACKLLNITADPMDWSPSNVQKWLL
                     WTEHQYRLPPMGKAFQELAGKELCAMSEEQFKQKSPLGGDVLHAHLDTWKSAAWMKER
                     TSPGATHYCASISEESWTDSEVDSSCSGQPTHLWQFLKELLLKPHSYGRFIRWLNEEK
                     GIFKIEDSAQVARLWGIRKNKPAMNYDKLSRSIRQYYEKGIIEKPDISQPLVYQFVHP
                     I"
BASE COUNT       468 a    654 c    571 g    302 t
ORIGIN


alignment scores:
            Quality:     9.00              Length:       9
              Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 DS-09-126-945b-2_copy_130_138 x AF071538   ..

 Align seq 1/1  to: AF071538  from: 1  to: 1894

        1 ValLeuLysAspIleGluThrAlaCys 9
          ||||||||||||||||||||||||||||
      803 GTGCTCAAGGACATCGAGACGGCCTGC 829

seq name: qb pr1:AB031549

seq documentation block:
LOCUS       AB031549      1898 bp    mRNA            PRI       20-JAN-2000
DEFINITION  Homo sapiens PSE mRNA for prostate ets, complete cds.
ACCESSION   AB031549
VERSION     AB031549.1  GI:6721497
KEYWORDS    prostate ets.
SOURCE      Homo sapiens adult male prostate epithelium prostate cancer
            cell line, LNCaP 3' cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Yamada,N., Tamai,Y., Miyamoto,H. and Nozaki,M.
  TITLE     Cloning and expression of the mouse Pse gene encoding a novel Ets
            family member
  JOURNAL   Gene 241 (2), 267-274 (2000)
  MEDLINE   20137504
REFERENCE   2  (bases 1 to 1898)
  AUTHORS   Nozaki,M.
  TITLE     Direct Submission
```

```
  JOURNAL    Submitted (25 AUG 2000) Yoshio [illegible] Department of Bacteriology, Osaka
             Nozaki, Osaka University, Institute for Microbial Diseases; 2-1,
             Yamadaoka, Suita, Osaka 565-0871, Japan
             (E-mail:mnozaki@biken.osaka-u.ac.jp, Tel:81 6-879 8340,
             Fax:81-6-879 8339)
FEATURES             Location/Qualifiers
     source          1. .1898
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="SP 3"
                     /cell_type="prostate cancer"
                     /chromosome="6"
                     /dev_stage="adult"
                     /map="6p21.2 21.2"
                     /sex="male"
                     /tissue_type="prostate epithelium"
                     /note="vector:pUC18"
     gene            406. .1413
                     /gene="PSE"
     CDS             406. .1413
                     /gene="PSE"
                     /codon_start=1
                     /product="prostate ets"
                     /protein_id="BAA89543.1"
                     /db_xref="GI:6721498"
                     /translation="MQSASPCTSSVSPSHLLIPPDIVSKYTSAAAACAVCQLEEKAAACAV
                     PSPPAIPEQGLSAEYLSYFDMLYPEDSSWAAKAPCASSPEFPPPEFPEQCPVIDSQAPA
                     GSLDLVPGGLELFEHSLEQVQSMVVGEVELKLELATCKLLNLIADFMEWSFSNVQKWLL
                     WTEHQYKLFPMQGAAPQELAGKELCAMSEEQFRQRSPLGQLDVLPAALPSWESAAWMKER
                     TSPGAIHYCASTSEESWTDSEVDSSCSGQLDIHLWQFLKELLLKPHSYCRFIEWLNKEE
                     GIFKLDSAQVAALWCLEKNRFAMNYLKLSKSLPQYEARLFLRDAPDLSQKLVVQLEVHP
                     I"
BASE COUNT       380 a    651 c    567 g    301 t
ORIGIN


alignment_scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_130_138 x AB031549  ..

 Align seq 1/1  to: AB031549  from: 1  to: 1898

        1 ValLeuLysAspIleGluThrAlaCys 9
          |||||||||||||||||||||||||||
    794 GTGYTCAAGGACATCGAGACGGGYTGC 819

seq_name: gb_pat1:AX048159

seq_documentation_block:
LOCUS        AX048159      3317 bp    DNA              PAT        15-DEC-2000
DEFINITION   Sequence 12 from Patent WO0070092.
ACCESSION    AX048159
VERSION      AX048159.1  GI:11876975
KEYWORDS     .
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo
REFERENCE    1  (bases 1 to 3317)
  AUTHORS    Kaufmann,J. and Xin,H.
  TITLE      Expression of ets-domain proteins in cancer
  JOURNAL    Patent: WO 0070092-A 12 23-NOV-2000;
             Chiron Corporation (US)
FEATURES             Location/Qualifiers
     source          1. .3317
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="GSEF encoding sequence with promoter (Figs. 2A 2B)"
     promoter        1. .1380
                     /note="Promoter"
     CDS             1915. .2922
                     /note="GSEF coding sequence"
                     /codon_start=1
                     /protein_id="CAC19004.1"
                     /db_xref="GI:11876976"
                     /translation="MQSASPCTSSVSPSHLLIPPDIVSKYTSAAAACAVCQLEEKAAACAV
                     PSPPAIPEQGLSAEYLSYFDMLYPEDSSWAAKAPCASSPEFPPPEFPEQCPVIDSQAPA
                     GSLDLVPGGLELFEHSLEQVQSMVVGEVELKLELATCKLLNLIADFMEWSFSNVQKWLL
                     WTEHQYKLFPMQGAAPQELAGKELCAMSEEQFRQRSPLGQLDVLPAALPSWESAAWMKER
                     TSPGAIHYCASTSEESWTDSEVDSSCSGQLDIHLWQFLKELLLKPHSYCRFIEWLNKEE
                     GIFKLDSAQVAALWCLEKNRFAMNYLKLSKSLPQYEARLFLRDAPDLSQKLVVQLEVHP
                     I"
BASE COUNT       711 a   1326 c    970 g    611 t
ORIGIN


alignment_scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_130_138 x AX048159  ..

 Align seq 1/1  to: AX048159  from: 1  to: 3317

        1 ValLeuLysAspIleGluThrAlaCys 9
          |||||||||||||||||||||||||||
    2... CTC ... AAGGACATCGAGACGGGCTGC 2228

seq_name: gb_pr6:AL157372

seq_documentation_block:
LOCUS        AL157372    150660 bp    DNA              PRI        29-NOV-2000
DEFINITION   Human DNA sequence from clone RP11-375E1 on chromosome 6, complete
             sequence.
ACCESSION    AL157372
VERSION      AL157372.18  GI:11494236
KEYWORDS     HTG.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 150660)
  AUTHORS    Skuce,C.
  TITLE      Direct Submission
  JOURNAL    Submitted (29 NOV 2000) Sanger Centre, Hinxton, Cambridgeshire,
             CB10 1SA, UK. Email enquiries: humquery@sanger.ac.uk Clone
             requests: clonerequest@sanger.ac.uk
COMMENT      On Nov 30, 2000 this sequence version replaced gi:11342772.
             During sequence assembly data is compared from overlapping clones.
             Where differences are found these are annotated as variations
             together with a note of the overlapping clone used. Note that the
             variation annotation may not be found in the sequence submission
             corresponding to the overlapping clone, as we submit sequences with
             only a small overlap as described above.
             This sequence has been finished according to sequence map criteria
             as follows.  An attempt is made to resolve all sequencing problems,
             such as compressions and repeats, but not necessarily within known
             annotated repeat sequence elements.  Where the sequence is
             ambiguous, there is an annotation using the 'unsure' feature key.
             The following abbreviations are used to associate primary accession
             numbers given in the feature table with their source databases:
             EMBL: EMBL; SW:, SWISSPROT; TR:, TREMBL; WP:, WORMPEP; Information
             on the WORMPEP database can be found at
             http://www.sanger.ac.uk/Projects/C_elegans/wormpep Each sequence
             was generated from part of bacterial clone contigs of human
             chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
             Group.  Further information can be found at
             http://www.sanger.ac.uk/HGP/Chr6
             RP11-375E1 is from the library RPCI-11.2 constructed at the Roswell
             Park Cancer Institute by the group of Pieter de Jong. For further
```

```
            details see http://bacpac.med.buffalo.edu/
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RPCI 475E1 it may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RPCI 475E1 is at 1 in this sequence. The
            true left end of clone RP3 391022 is at 150561 in this sequence.
            The true right end of clone RPCI 902L10 is at 113044 in this
            sequence.
FEATURES             Location/Qualifiers
     source          1..150660
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /chromosome "6"
                     /clone "RPCI 475E1"
                     /clone_lib "RPCI 11.2"
BASE COUNT     45470 a    38887 c    38740 g    27563 t
ORIGIN


alignment scores:
            Quality:     9.00            Length:       9
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_130_138 x AL157372/c=>

 Align seq 1/1  to reverse of: AL157372  from: 1  to: 150660

       1 ValLeuLysAspIleGluThrAlaCys 9
         ||||||||||||||||||||||||||||
  129454 GTGCTCAAGGACATCGAGACGGCCTGC 129427

seq_name: qb_pat2:109074

seq documentation block:
LOCUS        109074         885 bp                    PAT        02-DEC-1994
DEFINITION  Sequence 1 from Patent WO 8809814.
ACCESSION   109074
VERSION     109074.1  GI:588217
KEYWORDS    .
SOURCE      Unknown
  ORGANISM  Unknown
            Unclassified.
REFERENCE   1  (bases 1 to 885)
  AUTHORS   Chang,N.T., Chang,T.W., Fung,M.S-C., Fung,M.C., Gallo,R.C. and
            Wong-Staal,F.
  TITLE     EXPRESSION OF IMMUNOLOGICALLY ACTIVE PROTEINS OF HUMAN
            B LYMPHOTROPIC VIRUS
  JOURNAL   Patent: WO 8809814 A 1 15-DEC-1988;
FEATURES             Location/Qualifiers
     source          1..885
                     /organism="unknown"
BASE COUNT       465 a    214 c    124 g    182 t
ORIGIN


alignment scores:
            Quality:     7.00            Length:       7
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_130_138 x 109074  ..

 Align seq 1/1  to: 109074  from: 1  to: 885

       1 ValLeuLysAspIleGluThr 7
         |||||||||||||||||||
     286 GTGTTAAAAGACATAGAAACG 306

seq_name: qb_pl3:NEUM0S8
```

```
seq documentation block:
LOCUS        NEUM0S8        2018 bp    DNA             PLN       14-APR-2000
DEFINITION  Neurospora crassa DNA for mus-8, complete cds.
ACCESSION   D78372
VERSION     D78372.1  GI:1100728
KEYWORDS    mus-8.
SOURCE      Neurospora crassa DNA.
  ORGANISM  Neurospora crassa
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariales; Sordariaceae; Neurospora.
REFERENCE   1  (sites)
  AUTHORS   Soshi,I., Sakuraba,Y., Kafer,E. and Inoue,H.
  TITLE     The mus-8 gene of Neurospora crassa encodes a structural and
            functional homolog of the Rad6 protein of Saccharomyces cerevisiae
  JOURNAL   Curr. Genet. 30 (3), 224-231 (1996)
  MEDLINE   96337904
REFERENCE   2  (bases 1 to 2018)
  AUTHORS   Soshi,I., Sakuraba,Y., Kafer,E. and Inoue,H.
  TITLE     The mus-8 gene of Neurospora crassa encodes a structural and
            functional homolog of the Rad6 protein of yeast Saccharomyces
            cerevisiae
  JOURNAL   Unpublished (1996)
REFERENCE   3  (bases 1 to 2018)
  AUTHORS   Inoue,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-NOV-1995) to the DDBJ/EMBL/GenBank databases.
            Hirokazu Inoue, Saitama University, Department of Regulation
            Biology; Shimo-ohkubo 255, Urawa, Saitama 338, Japan
            (E-mail:inoue@sacs.sv.saitama-u.ac.jp, Tel:048-858-3413,
            Fax:048-858-3413)
FEATURES             Location/Qualifiers
     source          1..2018
                     /organism="Neurospora crassa"
                     /db_xref="taxon:5141"
     exon            697..799
                     /number=1
     5'UTR           697..757
     CDS             join(758..799,891..955,1039..1263,1337..1460)
                     /codon_start=1
                     /product="mus-8"
                     /protein_id="BAA11380.1"
                     /db_xref="GI:1100729"
                     /translation="MSTAAPPPELMPDPFRRMQIDPPAGVSASPVPPNVMPWNAVIIGFA
                     DIPPEDGTFRLVMHPEEQQYPNKPPSVKFISEMPHPNVYATGELCLDILQNRWSPTYDV
                     AAVLTSIQSLLNDPNPGSPANVEASNLYYQNPKEYHKPVRETVPKSWED"
     intron          800..890
     exon            891..955
                     /number=2
     intron          956..1038
     exon            1039..1263
                     /number=3
     intron          1264..1336
     exon            1337..1993
                     /note="one of the termination sites of exon4"
                     /number=4
     exon            1337..2018
                     /note="one of the termination sites of exon4"
     3'UTR           1461..1993
     3'UTR           1461..2018
     polyA_site      1993
                     /evidence=experimental
     polyA_site      2018
                     /evidence=experimental
BASE COUNT       453 a    527 c    501 g    537 t
ORIGIN


alignment scores:
            Quality:     7.00            Length:       7
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000
```

```
alignment_block:
 US-09-126-945B-2_COPY_130_138 x NEUMUS8_rev

 Align seq 1/1  to reverse of: NEUMUS8  from: 1  to: 2018

        1 ValLeuLysAspIleGluThr 7
          |||||||||||||||||||||
     1214 GTTCTGAAGGATATCGAGACA 1193

seq_name: db_pir:AN1291452

seq_documentation_block:
LOCUS       AN1291452     5000 bp    DNA             PLN       05-JAN-2001
DEFINITION  Aspergillus nidulans kinA gene for conventional kinesin, exons 1-5.
ACCESSION   AJ291452
VERSION     AJ291452.1  GI:12044814
KEYWORDS    kinA gene; kinesin.
SOURCE      Aspergillus nidulans.
  ORGANISM  Aspergillus nidulans
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Eurotiales; Trichocomaceae; Emericella.
REFERENCE   1  (bases 1 to 5000)
  AUTHORS   Requena,N. and Fischer,R.
  TITLE     Conventional kinesin of Aspergillus nidulans destabilises
            microtubules and thus controls nuclear distribution
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 5000)
  AUTHORS   Fischer,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-2000) Fischer R., Dept. of Biochemistry,
            Max-Planck-Institute for terrestrial Microbiology,
            Karl-von Frisch Str.  D-35043 Marburg, GERMANY
FEATURES             Location/Qualifiers
     source          1..5000
                     /organism="Aspergillus nidulans"
                     /db_xref="taxon:5072"
     mRNA            join(1167..1171,1238..1272,1336..1430,1501..3926,
                     4044..4266)
                     /gene "kinA"
                     /product "kinesin (KINA protein)"
     gene            1167..4266
                     /gene "kinA"
     exon            1167..1171
                     /gene "kinA"
                     /number=1
     CDS             join(1167..1171,1238..1272,1336..1430,1501..3926,
                     4044..4266)
                     /gene "kinA"
                     /function="motor protein  conventional kinesin"
                     /codon_start=1
                     /product="kinesin (KINA protein)"
                     /protein_id "CAC19836.1"
                     /db_xref="GI:12044815"
                     /translation "MASGTSNESNTLKVVAKFRQNKVELASGSQFTVEEPNDETCSI
                     NSKFASGSFTFDRVFDMDSFQTDVFNYSIAPTVETIFQYNETVFAYGQTGAGKSYT
                     MGSDIDDEVGKGIIPRIVFQIFASLTSPSNEFYTVFLSYMEIYMEKIRDLLVPKNGN
                     LPVHEPKSRGVYVKGLLFVYVSSVQEVYEVMRRGGNARAVAATNMNQESSRSHSIFV
                     TVTQKNLFTGSAKSGQLFLVDLAGSERVGKSGASGQTLEEAKFINKSLSALGDVMVNA
                     TDGKSTHIPYRDSKLTRILQESLGGNSRTTLIINCSPSSYNDAETVSTLRFGVRAKAI
                     KNKAKVNAELSPAELKQLLRRAQSQVISPENYISALESEVSSWRSGETVPREKWTPAR
                     NAEAVKGAKAEAKGPRPSTPSRLHLVSKSETTPFTGSRIGDKSSTPSTVLEKDEREEEL
                     RRENELQDQIAEKESHIAIIERCLEEAKQELEELENGAKCKLNEKLNAEVNELRMH
                     LEKVSYESKEAATIMDSLPEANSFLIAFLPVFQQLIVAVMFAFFIAAILKRFFFFA
                     EKMAKMMAGEDLGDEVESQNEKRLQNLIGVLSIRLIAAVCFVIAPDPLFLPVASLSF
                     TQGFIRQAELTVNDRSFLNVLQDSRKMELEQKLANLEQDYESLLARNLGEEDLAEIKE
                     RLEKVYIAKKDVFLAAASFLQQQLAKKDEFLIKLQQSLADSQSRQSINCAASKSLQQQ
                     IAEFDAMKKSLMSFLQNECDEVVELFLTGGASLYNRVLKSSNMEAVIEEMAILLRN
                     LEQLTHVQRQLVEQNSSLKKFVAIAPEFLTAFNFRLSTFAIILFSLFFLLAGHSF
                     AQLTAVKFRIPAAKQQSIKQLPGMDSNGGFSFAGSPLAKPLPQQQQNFPVAAVSTV
                     TGIKRTSWFFDRR"
     intron          1172..1237
                     /gene="kinA"
                     /number 1
     exon            1238..1272
                     /gene "kinA"
                     /number 2
     intron          1273..1335
                     /gene "kinA"
                     /number 2
     exon            1336..1430
                     /gene "kinA"
                     /number 3
     intron          1431..1500
                     /gene "kinA"
                     /number 3
     exon            1501..3926
                     /gene "kinA"
                     /number 4
     intron          3927..4043
                     /gene "kinA"
                     /number 4
     exon            4044..4266
                     /gene "kinA"
                     /number 5
BASE COUNT      1316 a   1269 c   1224 g   1189 t
ORIGIN


alignment_scores:
            Quality:    7.00              Length:       7
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_130_138 x AN1291452_rev

 Align seq 1/1  to reverse of: AN1291452  from: 1  to: 5000

        1 ValLeuLysAspIleGluThr 7
          |||||||||||||||||||||
       50 GTTTTGAAGGATATCGAGACA 30

seq_name: db_pir:AB042808

seq_documentation_block:
LOCUS       AB042808     7697 bp    RNA             VRL       06-JAN-2001
DEFINITION  Chiba virus genomic RNA, complete genome.
ACCESSION   AB042808
VERSION     AB042808.1  GI:11275471
KEYWORDS    .
SOURCE      Chiba virus (strain Ba/NIV/Chiba 447_1987/JP) genomic RNA.
  ORGANISM  Chiba virus
            Viruses; ssRNA positive-strand viruses, no DNA stage;
            Caliciviridae; Norwalk-like viruses.
REFERENCE   1  (sites)
  AUTHORS   Someya,Y., Takeda,N. and Miyamura,T.
  TITLE     Complete nucleotide sequence of the Chiba virus genome and
            functional expression of the 3C-like protease gene
  JOURNAL   Virology 278 (2), 490-500 (2000)
  MEDLINE   20569581
REFERENCE   2  (bases 1 to 7697)
  AUTHORS   Someya,Y., Takeda,N. and Miyamura,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-2000) Yuichi Someya, National Institute of
            Infectious Diseases, Department of Virology II; Toyama 1-23-1,
            Shinjuku, Tokyo 162-8640, Japan (E-mail:someya@nih.go.jp,
            Tel:81-3-3285-1111(ex.2523), Fax:81-3-3285-1161)
FEATURES             Location/Qualifiers
     source          1..7697
                     /organism="Chiba virus"
                     /strain="Ba/NIV/Chiba 447_1987/JP"
                     /db_xref="taxon:99565"
     gene            5..5362
                     /gene "ORF1"
     CDS             5..5362
                     /gene "ORF1"
```

```
                     /function "nonstructural"
                     /note "A polyprotein including RNA helicase, VPg, protease,
                     and RNA polymerase"
                     /codon_start 1
                     /product "polyprotein"
                     /protein_id "BAB18266.1"
                     /db_xref "GI:11275472"
                     /translation "MMMASKEVVAATAASAENSGNSETENPFEIKLKNSGNNKPI
                     KIENTHMALNLISRQSSPVPQQPPEDQRDEPPPKNVAELQQAMGWTDPEVDQNLPTWK
                     ELSQTEKQELLKNNSNWFDAGSLAPASTESAYVKNTFQQPFTDHQVKWSANNGVTELGVI
                     SLTTVPGPAWNLCDMPPPIDQRNNGPAREPLIGDMIEPYEGHIYHYAIYIQGKTVGVH
                     SPQAAFSIARTLIHPIAAWWRVCYIPTPDQGLNYDQLKELENEPWPYAAVTNNPCYEFC
                     CRILNIQDTWLFRRLVTSGGENHPAQDWSRDFDPGQQDSKLFTVRDAVIAATNCDVSK
                     PEKDLLNKLKPLNVLNLLSNCDWFEMGVVETTLLMELPGTFWNPPDVSSFIASLLPD
                     MHLQGPEDLAKDLLPLVLGGVGLAIGFTKDKVIKVMKSAVEGLEAATQLQQYGLELFA
                     VLKKYFEGGDQTEKTLKDTFIAVIDMEVTSTIAVIQLVRDKQAARIYMNTLDAEEKAA
                     RKLSVRHADDMVVSTNALTSKESLARSALAKAQAEMINKVKPVVVMMCGPGTGKTK
                     AAEYLAKRLANETRPGGKVGLVPREAVDDWDGYHGEEVMLWDDYGMSKIADDCNKLQA
                     IADTAPLTLNCDRTEKKGLQFVSDAIVITTNAIGPAPVIGFVNLGPVCRRVDFLVYCTA
                     PDVEQTRRTNDGDIGALKDCFKSDESHLKMELAPQGGFDNQGNTPFGKGIMRPTTLNR
                     LLTQAVALTMERQDEFQMQGQVYDFDADKVSAFTSLAKANGLGLISMASIGKKLRGVD
                     SVRGLKNALSGYTITPCSTKWQARVYDIPSDGPNVQTKRNTSAQTQRQQSTDTAALAL
                     TRLKAARAAAYAACTQGATTTLQMAGSAIVNRAVKRMPGAHSSTIALEGPPRERHC
                     RAHLAKAAGGPTGHDDVTDKYGLCKTERERGLSEEMNVEERATASLAGKNKGKTEKNGRG
                     RRANYNAFSRRGLSDEEYEEYKKTREEKNGNYSTQEYLEDRQRYTEEELAEVQAGGDRG
                     IGETEMETRHKTEYKSKKGRQNEKRQETGTVTGSDIRKKKETDWTFPKNDWAADDNREVD
                     YNEKLSFEAPPTLWSRVVRPGSGWCFWVSPTVFTTTHVIPTGVREFFGEPTESTAIR
                     RAGEFTQFRFSREVVRPDLTCMVLEECCPECVVCSTLTERDSGKELLPLAVEMCATASME
                     TQGRLVHGQSGMLLTGANAKGMDLGTLPGDCGAPVVYKRNNDWVVCGVHAAATKSGNT
                     VVCAVQAGFGETLLEGGDKGHYAGHETLKHSKGPALSTKTKPWRSTPPPLPPGVYEPA
                     YIGGRDPRVKGPSTQQGVLRQLKPFAEPRGRMPEPCLLEAAVETVTSMLEQTMDPPT
                     PWSFSDACQSLDKTTSSGHPYHKKRKNDDWNGTSEVGGFGQAAHANNMEFLGKSMKPL
                     YFAALKDELVKPEKVYQKIKKLLWGADGSIVIKAAKAPGPFCDATKPHVIKLPTKVG
                     MNSIEDGPMIYAEHAKYKNHFDADYSAWDSTQNKQTMTESFAIMCRLIASPELASVVA
                     KDLTAPSEMDVGDYTTRVKEGLPSGFPCTSQVNSINHWTTTCAMSEVIGSFDVTQS
                     QSYFSFYGDDETVSIDIDFDPARLIQVLEEYGLKFTPDDKSEDCPTTLRRQVDGLVFLR
                     RTISKDAAGPQGRLDRGSIERQLAWWTRGPNHDDPSETLITPHPQRKVQLTSLLGEASLH
                     GEKEYRKTSSKVTQETKTGGLEMYVPRWQAMFRWMRFHDLGLWTGDRNLLPEPVNDGG
                     V"
gene                 5446..6980
                     /gene "ORF2"
CDS                  5446..6980
                     /gene "ORF2"
                     /function "structural"
                     /note "ORF2 encodes a capsid protein which makes
                     nucleocapsid."
                     /codon_start 1
                     /product "capsid protein"
                     /protein_id "BAB18267.1"
                     /db_xref "GI:11275473"
                     /translation "MMMASKDAIPSALSATGAQGLVEEVNTAPDFEIIFVAGSSTALA
                     TAGQVNLIDPWTINNFVQAPQGEFTISPNNTPGDVLFDLQLGPHLNPFLSHLSQMYNG
                     WVGNMRVRVVLAGNAFTAGKVIISCVPPGFQSKTLSIAQATLFPHVIADVRTLDPVEV
                     PLEDVRNVLYHNNDFQPTMKLICMLYTPLRTGGAGGTDTFVVAGKVLSCEVELRPSNFN
                     FLVPPPTVEQFTFPEIVPNTPLEYLQNGKPTNPFEIDMSLCPSVQFQNVQPQNGETITVQ
                     PIGPTPVSVSQCQPPGGFFTSQQFVTNDTPLDQSPTMAPAAPAFASTPPDLGSCLEWHTE
                     MSKTLNLSDTQNPDTVTNCVKFFSWDVELKTTLKGVSTLMVSIVSPSQVDQKSYKIQN
                     GGANTNFWKTPDYGSSLAEASQLAPAVYTPQGNEVTVYGMASTPGPNQSQSPNLVPVS
                     LPQEYTTHPISEQAPTQGEAALLHYVDPDTNRNLGEFKLYPGGYLTCVPNSSSTGPQQ
                     LPLDGVFVFASWVSRFYQLKPVGTAGPARGRLGVRR"
gene                 6980..7606
                     /gene "ORF3"
CDS                  6980..7606
                     /gene "ORF3"
                     /function "unknown"
                     /note "ORF3 encodes a protein rich in basic amino acids,
                     whose function is unknown"
                     /codon_start 1
                     /product "basic protein"
                     /protein_id "BAB18268.1"
                     /db_xref "GI:11275474"
                     /translation "MAQATTGATAASAAGSAIGAGTQAGAEAATQAQRYQQDLTTQQN
                     SFNHDKEMLGYQMEMSNELLAKNENLRYSLTQAVGGLSSSDAAPAVAGAPVEPLVDWGC
                     VRVAAPQSSAITLRSGNEMAVPLEPAQPKQKPFLASEGYSNPAYDPVQLKTASWVQSQNSS
                     RGWGPYHRQATQTVWVTPPGRTSSSGVSIAPRGYFNTDRLPLFFANLRR"
BASE COUNT       2089 a    1787 c    1859 g    1962 t
ORIGIN


alignment scores:
                 quality:     7.00              length:        7
                   ratio:     1.000               gaps:        0
 Percent similarity: 100.000   Percent identity: 100.000

alignment block:
 US-09-126-945b-2_copy_130_138 x AB042808  ..

 Align seq 1/1   to: AB042808  from: 1  to: 7697

        2 LeuLysAspIleGluThrAla 8
          ||||:|||||||||||||||||
     1400 TTAAAGGATATTGAAACAGCC 1420

seq_name: qb_vt2:557540

seq_documentation_block:
LOCUS        S57540          8717 bp    DNA             VRL       21-JUN-1994
DEFINITION   ort RF1, 2 ort  LF2 [human herpesvirus 6, clone pZVH14, Genomic, 5
             genes, 8717 nt].
ACCESSION    S57540
VERSION      S57540.1  GI:235435
KEYWORDS     .
SOURCE       Human herpesvirus 6 clone pZVH14.
  ORGANISM   Human herpesvirus 6
             Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
             Betaherpesvirinae; Roseolovirus.
REFERENCE    1  (bases 1 to 8717)
  AUTHORS    Josephs,S.F., Abdel-Hadi,A.A., Salahuddin,S.Z., Lippolis,L.L.,
             Wong-Staal,F. and Gallo,R.C.
  TITLE      Identification of the human herpesvirus 6 glycoprotein B and
             putative large tegument protein genes
  JOURNAL    J. Virol. 65 (10), 5597-5604 (1991)
  MEDLINE    91374623
  REMARK     GenBank staff at the National Library of Medicine created this
             entry [NCBI qibbsq 57540] from the original journal article.
             This sequence comes from Fig.7.
FEATURES             location/qualifiers
     source          1..8717
                     /organism "Human herpesvirus 6"
                     /db_xref "taxon:10368"
                     /clone "pZVH14"
     gene            247..1254
                     /gene "ort RF1"
     CDS             247..1257
                     /gene "ort RF1"
                     /note "This sequence comes from Fig.7"
                     /codon_start 1
                     /protein_id "AAB19784.1"
                     /db_xref "GI:235436"
                     /translation "MDVCLYYPSSNKRLDLLFSTLNSLFLNLFLFLSPEL
                     DQQKIESLKIESKNLIAQPNGLYIPFLEKIPLSTLDEFLTENLFRHQLIYHIHQKI
                     TPQERNDLNKINQEFAIVIQQVIELEKSSCLPASLSQELDLLTERENKIGLNMEICKIY
                     IEILSNHMATPPTQPLINTFTINLRIVFIAHVFFICMENFSPLFLYYNKKKLLLEQQ
                     RAILLELRNEYSTLWKQLSDHIQCLLNVGLGLSLEKLYIRCGNEIGKQLLYENLIEKW
                     CNVFFLSISVSLFNNFSAHHLLLEGEDCKEVVQSLSLSAVSLSNLSIRPSFKLL
                     LR"
     gene            1202..2395
                     /gene "ort RF2"
     CDS             1202..2398
                     /gene "ort RF2"
                     /note "This sequence comes from Fig.7"
                     /codon_start 1
                     /protein_id "AAB19784.1"
                     /db_xref "GI:235437"
                     /translation "MNLYFFIALIHGSKHYYVKIYVIPMVSHLLNLLIEKLQSDCKLL
                     IIHACKLLLPSQHLLLHYMAWLYAFSINVQNIPLGIFFIVIESVIFFIADRINVMIHT
```

```
                     TYSPETNLGVSTLGNAYINTLQKYVNEWFSLLIANEIVLFLLFIPLGRAVLDFLWT
                     NTNLDNLEIYTTIGIDKIVITSPDSFTAIQKNLVKQHEEFEKSGLTIQSETTLIGML
                     RNITTSVSSSKEFITNFAIQKFTDVLPSFSQHVNEFVQSISMNLEFQPSNPLTLESL
                     FEDYFVPVESSNATLNTSIASPCTLEAANLAVCNNSETLERLEKDAVAVFKPFTTQLF
                     ESMKPTISLGKALMSTQKETIDHTTFLLIFPLDCKYNLVRHVIRQLNWYAAHAEAAQQDL
                     IIPLKF"
gene         1654. .1995
             /gene="orf RF3"
CDS          1654. .1998
             /gene="orf RF3"
             /note "This sequence comes from Fig.7"
             /codon_start=1
             /protein_id "AAB19785.1"
             /db_xref "GI:235438"
             /translation="MRQHIGDQMQHKLRQSIYTHNYRNRQNCHHFVELLLEEYMQKFSI
             STRRIKFITTNDPNFRNHTIRGDAIFKHNEELLQRAANQSSTTKTHHHFCTTHLPTEKR
             NIPVYFHEPRKM"
gene         2416. .8646
             /gene="orf RF5"
CDS          2416. .8649
             /gene="orf RF5"
             /note "large tegument protein; this sequence comes from
             Fig.7"
             /codon_start=1
             /protein_id "AAB19786.1"
             /db_xref "GI:235439"
             /translation "MFLTPSPINLNSSFSTYRAGEQVMCNPSLVRYVIRHVVLNCNPSCQ
             AGFIDAIMFPCYHLITACPLALMCNNPTQGVRLFTLIFKFFIRSSVCVTQHSLAFFPDG
             GFLDTQKIDNFVYLGIIDLILYGKSPNGSTPAVFLFCVSFFALFLLNNILVLFVSIPVT
             EREAIAAIYIGQNIEEVYELLITHGTEGFYYDASLIFFTEISNLSLSSHDAELLILKI
             YKDPDIAIALDKFSSIFIHFIKFIDDGSQQNGVADKLILLSPAPHPKKKKNGSLSLLE
             LNDKKKKDTASLIYYATEVDLIPSFYELRSQFQSLFRDLKSFPIMKSNFNWLIYLQDS
             PINPNQPFATPFLWNRVFHLIQQIIDVFVGVGSTNDDSSKQQQQTIFINYLIPFKDFS
             EVFNEALAACQENNLDIIFIYNNYLDKTTTFRILERILLSKFLAIVDNDHKKHYEWVK
             SWTTQMFQGMPKFIGTFNYIFAIVGHNPVKHEHEFIGINEAEEYKVAVLLNEKKKEI
             QEAIEREKNSFAQLSNFIDKLGETPALPTIESENVHKVHTSDITEGIVPREITTESIELP
             NISTLNNTQQISLDKQLNEKLTNITIHTLTNKFTKIVQDNYNNIAAGFMPVTELNCLFA
             YLVNLYFNIEVLKHSGINTNTVIIQFVFKLYDNTQFLPFCTSHPNINNISNPTISIPK
             MFVDEYFNSQKPSDPASFIIAAIFSIGAQPSKNEIVVNTFMIESGLIIFDERMLIITEN
             KQIAAITKQTLGDQFLTPIYDFLHHLSAYNLPNITTVKNLHLHFILEQRPDIAMILHD
             KIQSILDIYVIDMLNDITVPEQIFSIVLFLVDLFPNSTEKIALFESVLILRQLAKKCA
             NLFTLDFFDDLAQFLTNSFQIQNMMPQHFKEIFTLMCHLKFLYSQKITIIDPFNWI
             QRAKIVVITSPFFLTAFLATAPIFHAIQCPPGIPALQSMELSMPLIALNDKFFSI
             TIRSTLEKRLNDILIIKDGQFSSLFTMHLNELTFLKQLQDNNVIIHFTHALLPVLK
             DIETTISKITSDVIEKILNFLGMSKEEGKYIFLLSLSAFKAIEFCIEDVKIE
             IDQMQFSTIFLKIATSANFHTFISRIIVSQLFDNCYLILTIELEELLNRMKLQLSEY
             ALAFKKILLSSQQAKIDKIYIVSNLMLFLKKNQLKFCALAKALIVEVNNFFNQLIE
             LLNHFNAHLKAKMDQNHIIKLSFDFFWTAFVSDSKLYLPDSINIKIQDFISDPFKVIS
             QLMNKAINEMPYIQAFTLFWLIKLIHDINKFCLSAISFFGKFAIPFNYAAIPIQEYQ
             INTKYVEIFNKVIGNFIVENIENIPKLIKLLQLVPKPVAGCQEQQYQTLMNKLLSSPT
             SMQQTYFKFQLIFEYIIIVNNVAFILAENLPSNLNKIKLLIKSVFTLIKSQVFLFIT
             QENDLLSDSLNIENYINGLRALLNFLTAAQNYIQNTLLKQWAVDQQQNFIFIDYSVAN
             VKFISDLYAKLRIEIEKQVFTQVNSVFGTHLIVDLIGVPIQQGHNIFNNAIVKFFSLNY
             KQIRVPFFDIPPLVSSQYKLESVCKSFIMIIQQPWENIIIFLFGPYLPQGIQNFPQLLI
             PIVNLKLFIYCITQAWTASPDSIVCTATPTIPKPTTSTPPPVSVCTQRSTFCVV
             INSLAKSIDKNSLYNIFVVSHNIFFFMHEMPNTCIFLFWLIAFLMFITFQIFTIY
             LCPKNPKFFIYIIAFLIIPYFFMNYIWIQYFAAFIFQFSYQNLIFIIISEYVHQNSIT
             TSSVTPHEPDTIKSGERLISFIVHSAQNFPTLIRIQAQEYVFDYILVSFLICYEMIF
             AMYIDITEKTYLLCMKHLENVLHPIKIFQSVLPAKIFDIALIEKQSWKNIVEHSLFSV
             QLDKIVSYLNHINPAIPNIPLLFENYLNPVVNVYIPPMLCNFFFVADYIFNIFIHFIVL
             EYEATNLISPHLYDKIIDIINQLPNNTESTPPPPIPLNLLIENLSLFFFKKIPAVIC
             LMPKSQSIILSTDINVPETSPDVKANIASAAIKDVILAKEEINEESESINIIISELKS
             MYL"
CDS          complement(4719. .5027)
             /gene="orf  LF2"
             /note="This sequence comes from Fig.7"
             /codon_start=1
             /protein_id="AAB19787.1"
             /db_xref="GI:235440"
             /translation "MNWAFSLNSCSAVIFLARPFAMPEAVPLLPCSVPSVLILIFLCF
             NRFVFNVCSGTVMSLSMSSTYMSSMFIILSCSVIAISGLSLMKWFCKFFIVVVLCRL
             "
gene         4722. .5027
             /gene="orf  LF2"
BASE COUNT       3097 a    2220 c    1261 g    2139 t
ORIGIN


Alignment_scores:
              Quality:     7.00              Length:       7
                Ratio:    1.000               Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

Alignment_block:
 US-09-126-945B-2_COPY_130_138 x S57540  ..

 Align set 1/1   to: S57540   from: 1   to: 8717

       1 ValLeuGlyAspLeuGlnThr 7
         ||||.|||||.|||||||||
    5497 GTGTCAAAAGACAACAGAAACC 5517

seq_name: qb_vi2:HH6IMMEARH

seq_documentation_block:
LOCUS        HH6IMMEARH   38491 bp    DNA             VRL       25-JUL-2000
DEFINITION   Human herpesvirus 6 immediate-early gene homologs.
ACCESSION    L25528
VERSION      L25528.1  GI:451932
KEYWORDS     DNA polymerase processivity factor; G-protein
             coupled receptor; capsid protein; glycoprotein; immediate early
             gene; ribonucleotide reductase large subunit; tegument protein;
             virion protein.
SOURCE       Human herpesvirus 6.
  ORGANISM   Human herpesvirus 6
             Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
             Betaherpesvirinae; Roseolovirus.
REFERENCE    1  (bases 1 to 38491)
  AUTHORS    Nicholas,J. and Martin,M.E.
  TITLE      Nucleotide sequence analysis of a 38.5-kilobase-pair region of the
             genome of human herpesvirus 6 encoding human cytomegalovirus
             immediate early gene homologs and transactivation functions
  JOURNAL    J. Virol. 68 (2), 597-610 (1994)
  MEDLINE    94118404
   PUBMED    8289364
FEATURE                  Location/Qualifiers
     source              1. .38491
                         /organism "Human herpesvirus 6"
                         /strain="U1102"
                         /db_xref="taxon:10368"
     gene                <1. .838
                         /gene "PIRF0"
     CDS                  1. .295
                         /note "PIRF0"
                         /codon_start=2
                         /protein_id="AAA16715.1"
                         /db_xref "GI:451933"
                         /translation "EFFLHFTLKSFFSLHLFGSTIAILFYVLRMLGNFGLVI
                         AFLSNGWHDMLRLFLNGLHHGALADSIASPKFLAAHGPIVFDPVSPS"
     gene                complement(158. .2814)
                         /gene "PILF1"
     polyA_signal        complement(158. .163)
                         /gene "PILF1"
     CDS                 complement(199. .2814)
                         /gene "PILF1"
                         /codon_start=1
                         /product "virion protein"
                         /protein_id="AAA16716.1"
                         /db_xref "GI:451934"
                         /translation "MNEGQRHFFTAWLLKQKVIESTLTSNGLKLLNVTLKHFHVI
                         NSFVVFLCLVCLLKLLNLAVWLMYQVVSKEELAMNAWSPVTLLSVVNEYLNS
                         FGLNKSIFFNMFTNFFFSQFSDINFAFIFLGNFIKWSSNVAIDIFYVNLIAFISSE
                         TNNDLGAIFKNMCWYIVYNFNGWDLNQYLISINFLVLFALLWFFLIVA
                         MSGVIVTLQNHLSGHLRKHLNFNFIVYSHRVLQTFTGTGQKVESFLKTTSKYLTIKSSLE
                         SHSASKAFSMSPIGPNSLMDFVPLPGDTHSNLTLPSMSTDIKKSSLDVARLKKSNSRS
                         LDSFLKMQKQPKFLFLLSSVLNAGEKILLKEAILGGENVKAITPASSVSLMSGVFSPSS
```

FTSTNLDDLPLSSFLVTNGLSSLECDLNYEDGPPSSSTNVEFELNAGLNGNVELSL
KTSTGSPSSSGTTDVVNGFANLNLDGNKGNVSFDWGRNTAANADFLDFVHKEVPEQGF
PFVLNNSDVAGSFAKHTTYSTETGVSFKNVFLTKDAGKDGRAGRQSDLFFKSCLFGR
NDKATMHSKEVTGDSGDATELVGARNSPALKKTKQANDEFAGLGNKKNDRDVLRGGKGN
SKDLDEGGNAKKKLMSGKGNDGKLMTRNGSFLGGLTMGVAKNAGDEQYTQAGTAQRVN
NLLSQFTNLSLSFFGTEFTLQNQPLSSKLALTLNEAVGRLSLLANVERTLTVSNPGFM
FFNFRLQNRLISVGSTFELPGAFLSRELSASGTEDALLKLDPNGELELTGALLKVSV
GETETSDLKVGQDESFVPAQLMKVELTPEEKDDLTEQMVLPTRQDGETDENTVSGTGVA
ESLDIEAKGESATAS"

polyA_signal 838..843
/gene="EFRF0"

gene 4023..4074
/gene="EFRF1"

CDS 4023..4045
/gene="EFRF1"
/note="hypothetical"
/codon_start=1
/product="G protein coupled receptor"
/protein_id="AAA16717.1"
/db_xref="GI:451938"
/translation="MNFEFRGNASVTSIPGTLKIAKIMFSAVTGLFLLTSTPMLITVVF
IMILYHRVAKHNATSFYVITLAFSDLFVLMWVFFLMIVNRKVLFGLNKFFVCQLVYFTYH
AVCSYSISMLALIALTIERYKLIHRPKEFISSSSTKNIGIIILASVSMSAFTALPFVFT
NGMKKIGRCVVYYSSKKAYFTLAVFLVFSFIWVLFPMVFSFFYVTICKAIRQVTPP
KYKKTLFFTRILLLSPLLIQIPYIAILICKIAFLYMPQNICFWLARVFTLQLLIPLMP
QVHCFSNFLVYAFTGGELKNKFIACFQSFFFLIGTSIQFFVSTVASFHLQNSFSASV
EKNQPL"

polyA_signal 4069..4074
/gene="EFRF1"

gene 4128..5042
/gene="EFRF1"

CDS 4128..4451
/gene="EFRF1"
/codon_start=1
/protein_id="AAA16718.1"
/db_xref="GI:451936"
/translation="MAHAFFFVPPFLLVIGNCLNSLTLMPFIFKIAFAFLICFSLHNC
DYIASKLGIDFFFISPLSVLGFIVLLLGYVIGNNGLAIVYFIFFLLCTFLHNC
LSPFL"

gene 4507..6442
/gene="EFRF2"

CDS 4507..6378
/gene="EFRF2"
/note="similar to herpes cytomegalovirus UL29 gene family"
/codon_start=1
/protein_id="AAA16719.1"
/db_xref="GI:451937"
/translation="YIPYEIFKVVEITPMVTHAMEEPKTFNITTFVLDENCNFTPEVL
SPANAKYSKDVLGQSYNIVCTANSSIPMLIGSAFQSMVLLKEEFSPARQCFEEVLP
LGVHINREFYAGKKQVKHMCTMMKSLFDTEEAMRKLDKALMGLFVDARGSSYMPLIALS
LRENGLPDSKFTKAVKTINTTFNCTHNWTDAVTPQVAQFSAVNYFYKTPHVIEFFAV
DVYSENLKDLSIGVNKKFTLLFTSSDDGYLSHIFNDMLFLTSTWNMIYNCKVQTERLL
TWIKYKFINSIVELCVLVGFQLPDLKTFTTDLAALCSNMNLVDPDKDLFSDYKLTLEKL
FETSTPATKANICTLPTFTESHLTEEFDVLKPSNIPDDLNYLLLKSNDSPDEYDPEL
PTQVDPGKVDNVLTDSNFFNVTADNAFSSTAIMFTSYDKTLDVIENNVTQVLVEVEMQSL
SAVVYCAVANKYQDSLPQVTKRLNQNFLFAASEAASSHSSSFSFPYSFTQSSESFTSI
LRQRAPTRSNSRSSSVSPSQDDRNRSHYSDETNTSDYSYTMADLDLEDEEPMEDHPHS
PQSASSNNSMSRRGRALQNESKRKTTPMAPTSPARSQNNARLKKKTRKSTPTNLEEL"

polyA_signal 5037..5042
/gene="EFRF1"

polyA_signal complement(6390..6395)
/gene="EFLF3"

polyA_signal 6437..6442
/gene="EFRF2"

gene complement(6893..7294)
/gene="EFLF3"

CDS complement(6893..7294)
/gene="EFLF3"
/codon_start=1
/protein_id="AAA16720.1"
/db_xref="GI:451938"
/translation="TSICTYNTESYKAERNKALAVIKMDVWKEQELQICRELCTLFVL
MSLLSNMFSKIFLVYVKYLFKMDFSIMYKYLLPALLLSMLVLKSLVLFMLFLLKMWFDL

/gene="EFLF3"

gene complement(7427..8449)
/gene="EFLF2"

gene complement(7427..8600)
/gene="EFLF1"

polyA_signal complement(7427..7432)
/gene="EFLF1"

CDS complement(7492..8449)
/gene="EFLF2"
/note="similar to human cytomegalovirus US22 gene family; putative"
/codon_start=1
/product="immediate-early protein"
/protein_id="AAA16721.1"
/db_xref="GI:451939"
/translation="VSVLFPVSSVVCHSPSCTLNFVSFLFLKQGSLSLCCPGWQWCGY
SVWGFELKDVCLLKLLVFAGFGEKFYGLDFVNFLYLLAFLFEFAVLGLKRYFLVYK
KKFLFFLLKLPSVTQGSCLFAQPNLWKLAFIVYKNGSHDYSMKSSFLGNNGSVFVLFLK
KKSFSPMFWNFFSLVYVSMPYFSLVFGSFGKLLGLPGLFVAMMLSSAAIYGLFLCL
CCVFPIADGFTCFLFPCCVFGYFFYFLHRKTNIVQLFLFSTHLCSGFTCSNNYI
YDSSSEDDE"

CDS complement(8181..8600)
/gene="EFLF1"
/note="similar to human cytomegalovirus US22 gene family; putative"
/codon_start=1
/product="immediate early protein"
/protein_id="AAA16722.1"
/db_xref="GI:451940"
/translation="EFRNVLMAFFFFLFVVKIGYLFALMKLNSVLKLAVFLSCLGYVE
GFLCCLLFFNWLPFSPCLVISLCCLLFFLFSAARLGFMCFYGCFVSVFWCVLLVFVVR
IKLFFLLNKVRYYFAAKKDCNYLDMFLSGERRLVMCCA"

gene 8958..9748
/gene="EFRF3"

CDS 8958..9248
/gene="EFRF3"
/codon_start=1
/protein_id="AAA16723.1"
/db_xref="GI:451941"
/translation="NLGYKKAINFGNLFNLKSLFGKFLGLGLIAALLLTMLALNLNFFFI
YFNGRNNFFLTFGFRQLHKRGFCNFFLCLFMVFNLFNVYFLLFLFLPYLCFHC"

gene complement(9643..10728)
/gene="EJLF5"

gene complement(9643..10644)
/gene="EJLF6"

gene complement(9643..12120)
/gene="EJLF4"

polyA_signal complement(9643..9648)
/gene="EJLF5"

polyA_signal 9733..9738
/gene="EFRF3"

CDS complement(9741..10644)
/gene="EJLF6"
/note="glycoprotein; putative"
/codon_start=1
/product="immediate early protein"
/protein_id="AAA16724.1"
/db_xref="GI:451942"
/translation="TLNVMKKLIMESLSVYTFVMGVCFLSNFLCLQKIVLLQEHKLKS
QCSLFCYVNCVLAGNGAFVSVKLSYLLNLAMLFNPLFAMLVGNLLSLFKFAFLKVLI
NYYTRGVMLKALIAQKLFNAANLSGTVNCWLIRELAGSVMLLFVGLLSVLNSSLEIN
ASPWISLCLQVANCLLNLLNLPVSLDIYLLSKLCPQLSSLFLYEKKVAHLKYFSLLCKNDI
KFKKFLLRNTWKSAWSNWYKYKLIKLLLFGSGDYLNFSLLTYSMSAAGLFFLAGGAFIM
LLLLCCLSMLTRKHIVKDLGY"

CDS complement(10513..10728)
/gene="EJLF5"
/codon_start=1
/protein_id="AAA16725.1"
/db_xref="GI:451943"
/translation="WVCTLHFCYLFCCLPGFCFFLGFNCKFFFVNFKCYIFVNYGVL
VSLYLRNCSLLYIFFHMRAKNSLNSGI"

CDS complement(10882..12120)
/gene="EJLF4"

```
                     /note "putative"
                     /codon_start 1
                     /product="immediate-early protein"

alignment_scores:
             Quality:     7.00              Length:       7
               Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_130_138 x HH6IMMEARH  ..

 Align seq 1/1  to: HH6IMMEARH  from: 1  to: 38491

       1 ValLeuLysAspIleGluThr 7
         |||||||||||||||||||||
   29465 GTGTTAAAAGACATAGAAACG 29485

Seq_name: gb_htg18:AC087855

seq_documentation_block:
LOCUS       AC087855     61958 bp    DNA              HTG       01-FEB-2001
DEFINITION  Homo sapiens chromosome 8 clone RP11-431M3 map q, 1 WORKING DRAFT SEQUENCE
            SAMPLING.
ACCESSION   AC087855
VERSION     AC087855.1  GI:12643077
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 61958)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-431M3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 61958)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Holmes,W., Iliev,I., Johnson,P.,
            Jones,C., Karatas,A., LaRoeque,K., Lamazares,R., Landers,I.,
            Lehoczky,J., Levine,R., Liu,C., Maclean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenqa,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Risc,C., Rodov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schanor,S., Schupback,P., Seamon,S., Savery,P.,
            Sousnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
            Travers,M., Travis,N., Triqilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zemback,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-FEB-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker.
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ---------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www.seq.wi.mit.edu
                Contact: sequence-submissions@genome.wi.mit.edu
            ---------- Project Information
                Center project name: LL1715
                Center clone name: 431_M_3
            ---------- Physical Assembly
            * NOTE: This record contains 75 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *         1      735: contig of 735 bp in length
            *       736  835: gap of      100 bp
            *       836     1568: contig of 734 bp in length
            *      1569  1668: gap of      100 bp
            *      1669     2408: contig of 740 bp in length
            *      2409  2508: gap of      100 bp
            *      2509     3251: contig of 743 bp in length
            *      3252  3351: gap of      100 bp
            *      3352     4095: contig of 744 bp in length
            *      4096  4195: gap of      100 bp
            *      4196     4948: contig of 753 bp in length
            *      4949  5048: gap of      100 bp
            *      5049     5773: contig of 725 bp in length
            *      5774  5873: gap of      100 bp
            *      5874     6601: contig of 728 bp in length
            *      6602  6701: gap of      100 bp
            *      6702     7435: contig of 734 bp in length
            *      7436  7535: gap of      100 bp
            *      7536     8265: contig of 730 bp in length
            *      8266  8365: gap of      100 bp
            *      8366     9092: contig of 727 bp in length
            *      9093  9192: gap of      100 bp
            *      9193     9914: contig of 722 bp in length
            *      9915 10014: gap of      100 bp
            *     10015    10746: contig of 732 bp in length
            *     10747 10846: gap of      100 bp
            *     10847    11595: contig of 749 bp in length
            *     11596 11695: gap of      100 bp
            *     11696    12427: contig of 732 bp in length
            *     12428 12527: gap of      100 bp
            *     12528    13243: contig of 716 bp in length
            *     13244 13343: gap of      100 bp
            *     13344    14079: contig of 736 bp in length
            *     14080 14179: gap of      100 bp
            *     14180    14903: contig of 724 bp in length
            *     14904 15003: gap of      100 bp
            *     15004    15736: contig of 733 bp in length
            *     15737 15836: gap of      100 bp
            *     15837    16565: contig of 729 bp in length
            *     16566 16665: gap of      100 bp
            *     16666    17393: contig of 728 bp in length
            *     17394 17493: gap of      100 bp
            *     17494    18221: contig of 728 bp in length
            *     18222 18321: gap of      100 bp
            *     18322    19053: contig of 732 bp in length
            *     19054 19153: gap of      100 bp
            *     19154    19868: contig of 715 bp in length
            *     19869 19968: gap of      100 bp
            *     19969    20698: contig of 730 bp in length
            *     20699 20798: gap of      100 bp
            *     20799    21534: contig of 736 bp in length
            *     21535 21634: gap of      100 bp
            *     21635    22340: contig of 706 bp in length
            *     22341 22440: gap of      100 bp
            *     22441    23187: contig of 747 bp in length
            *     23188 23287: gap of      100 bp
            *     23288    23985: contig of 698 bp in length
            *     23986 24085: gap of      100 bp
            *     24086    24811: contig of 726 bp in length
            *     24812 24911: gap of      100 bp
            *     24912    25645: contig of 734 bp in length
            *     25646 25745: gap of      100 bp
            *     25746    26470: contig of 725 bp in length
```

```
*     26471 26570: gap of       100 bp
*     26571     27312: contig of 742 bp in length
*     27313 27412: gap of       100 bp
*     27413     28145: contig of 733 bp in length
*     28146 28245: gap of       100 bp
*     28246     28970: contig of 725 bp in length
*     28971 29070: gap of       100 bp
*     29071     29806: contig of 736 bp in length
*     29807 29906: gap of       100 bp
*     29907     30633: contig of 727 bp in length
*     30634 30733: gap of       100 bp
*     30734     31473: contig of 740 bp in length
*     31474 31573: gap of       100 bp
*     31574     32296: contig of 723 bp in length
*     32297 32396: gap of       100 bp
*     32397     33151: contig of 755 bp in length
*     33152 33251: gap of       100 bp
*     33252     33993: contig of 742 bp in length
*     33994 34093: gap of       100 bp
*     34094     34823: contig of 730 bp in length
*     34824 34923: gap of       100 bp
*     34924     35626: contig of 703 bp in length
*     35627 35726: gap of       100 bp
*     35727     36452: contig of 726 bp in length
*     36453 36552: gap of       100 bp
*     36553     37286: contig of 734 bp in length
*     37287 37386: gap of       100 bp
*     37387     38119: contig of 733 bp in length
*     38120 38219: gap of       100 bp
*     38220     38940: contig of 721 bp in length
*     38941 39040: gap of       100 bp
*     39041     39806: contig of 766 bp in length
*     39807 39906: gap of       100 bp
*     39907     40627: contig of 721 bp in length
*     40628 40727: gap of       100 bp
*     40728     41467: contig of 740 bp in length
*     41468 41567: gap of       100 bp
*     41568     42298: contig of 731 bp in length
*     42299 42398: gap of       100 bp
*     42399     43115: contig of 717 bp in length
*     43116 43215: gap of       100 bp
*     43216     43829: contig of 614 bp in length
*     43830 43929: gap of       100 bp
*     43930     44658: contig of 729 bp in length
*     44659 44758: gap of       100 bp
*     44759     45488: contig of 730 bp in length
*     45489 45588: gap of       100 bp
*     45589     46305: contig of 717 bp in length
*     46306 46405: gap of       100 bp
*     46406     47136: contig of 731 bp in length
*     47137 47236: gap of       100 bp
*     47237     47960: contig of 724 bp in length
*     47961 48060: gap of       100 bp
*     48061     48786: contig of 726 bp in length
*     48787 48886: gap of       100 bp
*     48887     49611: contig of 725 bp in length
*     49612 49711: gap of       100 bp
*     49712     50447: contig of 736 bp in length
*     50448 50547: gap of       100 bp
*     50548     51261: contig of 714 bp in length
*     51262 51361: gap of       100 bp
*     51362     52086: contig of 725 bp in length
*     52087 52186: gap of       100 bp
*     52187     52917: contig of 731 bp in length
*     52918 53017: gap of       100 bp
*     53018     53746: contig of 729 bp in length
*     53747 53846: gap of       100 bp
*     53847     54576: contig of 730 bp in length
*     54577 54676: gap of       100 bp
*     54677     55403: contig of 727 bp in length
*     55404 55503: gap of       100 bp
*     55504     56232: contig of 729 bp in length
*     56233 56332: gap of       100 bp
*     56333     57073: contig of 741 bp in length
*     57074 57173: gap of       100 bp

alignment scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY 130-138 x AC087855    .

 Align seq 1/1   to: AC087855  from: 1  to: 61958

       2 LeuLysAspIleGluThrAla 8
         ||||||||||||||||||||||
   44652 TTGAAACATATAGACACAGCC 44672

seq name: gb_htg2:AC010861

Seq documentation block:
LOCUS       AC010861     63369 bp    DNA             HTG       10-DEC-2000
DEFINITION  Homo sapiens clone RPTI 4D5, LOW PASS SEQUENCE SAMPLING.
ACCESSION   AC010861
VERSION     AC010861.3  GI:11612363
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 63369)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RPTI 4D5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 63369)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boqvslavskiy,L., Bookhalter,B.,
            Brown,A., Castle,A., Colanqelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeAreiiano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHuqh,W., Forrest,C., Funke,R., Gaqe,D.,
            Galaqan,J., Gardyna,S., Grant,G., Haqos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lien,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,F., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-SEP-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 10, 2000 this sequence version replaced gi:10280836.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www.seq.wi.mit.edu
                Contact: sequence submissions@genome.wi.mit.edu
                -    Project Information
                Center project name: L2717
                Center clone name: 4_D_5
                -
            * NOTE: This record contains 77 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
```

```
* the record is updated, the accession number will
* be preserved.
*        1      747: contig of 747 bp in length
*      748  847: gap of      100 bp
*      848     1539: contig of 692 bp in length
*     1540  1639: gap of      100 bp
*     1640     2253: contig of 614 bp in length
*     2254  2353: gap of      100 bp
*     2354     3034: contig of 681 bp in length
*     3035  3134: gap of      100 bp
*     3135     3861: contig of 727 bp in length
*     3862  3961: gap of      100 bp
*     3962     4681: contig of 720 bp in length
*     4682  4781: gap of      100 bp
*     4782     5513: contig of 732 bp in length
*     5514  5613: gap of      100 bp
*     5614     6346: contig of 733 bp in length
*     6347  6446: gap of      100 bp
*     6447     7161: contig of 715 bp in length
*     7162  7261: gap of      100 bp
*     7262     8017: contig of 756 bp in length
*     8018  8117: gap of      100 bp
*     8118     8874: contig of 757 bp in length
*     8875  8974: gap of      100 bp
*     8975     9712: contig of 738 bp in length
*     9713  9812: gap of      100 bp
*     9813    10503: contig of 691 bp in length
*    10504 10603: gap of      100 bp
*    10604    11289: contig of 686 bp in length
*    11290 11389: gap of      100 bp
*    11390    12115: contig of 726 bp in length
*    12116 12215: gap of      100 bp
*    12216    12933: contig of 718 bp in length
*    12934 13033: gap of      100 bp
*    13034    13770: contig of 737 bp in length
*    13771 13870: gap of      100 bp
*    13871    14623: contig of 753 bp in length
*    14624 14723: gap of      100 bp
*    14724    15474: contig of 751 bp in length
*    15475 15574: gap of      100 bp
*    15575    16305: contig of 731 bp in length
*    16306 16405: gap of      100 bp
*    16406    17101: contig of 696 bp in length
*    17102 17201: gap of      100 bp
*    17202    17891: contig of 690 bp in length
*    17892 17991: gap of      100 bp
*    17992    18705: contig of 714 bp in length
*    18706 18805: gap of      100 bp
*    18806    19501: contig of 696 bp in length
*    19502 19601: gap of      100 bp
*    19602    20329: contig of 728 bp in length
*    20330 20429: gap of      100 bp
*    20430    21164: contig of 735 bp in length
*    21165 21264: gap of      100 bp
*    21265    22007: contig of 743 bp in length
*    22008 22107: gap of      100 bp
*    22108    22846: contig of 739 bp in length
*    22847 22946: gap of      100 bp
*    22947    23681: contig of 735 bp in length
*    23682 23781: gap of      100 bp
*    23782    24527: contig of 746 bp in length
*    24528 24627: gap of      100 bp
*    24628    25311: contig of 684 bp in length
*    25312 25411: gap of      100 bp
*    25412    26090: contig of 679 bp in length
*    26091 26190: gap of      100 bp
*    26191    26904: contig of 714 bp in length
*    26905 27004: gap of      100 bp
*    27005    27712: contig of 708 bp in length
*    27713 27812: gap of      100 bp
*    27813    28551: contig of 739 bp in length
*    28552 28651: gap of      100 bp
*    28652    29376: contig of 725 bp in length
*    29377 29476: gap of      100 bp
*    29477    30220: contig of 744 bp in length
*    30221 30320: gap of      100 bp
*    30321    31067: contig of 747 bp in length
*    31068 31167: gap of      100 bp
*    31168    31934: contig of 767 bp in length
*    31935 32034: gap of      100 bp
*    32035    32789: contig of 755 bp in length
*    32790 32889: gap of      100 bp
*    32890    33598: contig of 709 bp in length
*    33599 33698: gap of      100 bp
*    33699    34405: contig of 707 bp in length
*    34406 34505: gap of      100 bp
*    34506    35200: contig of 695 bp in length
*    35201 35300: gap of      100 bp
*    35301    36021: contig of 721 bp in length
*    36022 36121: gap of      100 bp
*    36122    36894: contig of 773 bp in length
*    36895 36994: gap of      100 bp
*    36995    37724: contig of 730 bp in length
*    37725 37824: gap of      100 bp
*    37825    38550: contig of 726 bp in length
*    38551 38650: gap of      100 bp
*    38651    39420: contig of 770 bp in length
*    39421 39520: gap of      100 bp
*    39521    40233: contig of 713 bp in length
*    40234 40333: gap of      100 bp
*    40334    41062: contig of 729 bp in length
*    41063 41162: gap of      100 bp
*    41163    41869: contig of 707 bp in length
*    41870 41969: gap of      100 bp
*    41970    42651: contig of 682 bp in length
*    42652 42751: gap of      100 bp
*    42752    43441: contig of 690 bp in length
*    43442 43541: gap of      100 bp
*    43542    44252: contig of 711 bp in length
*    44253 44352: gap of      100 bp
*    44353    45074: contig of 722 bp in length
*    45075 45174: gap of      100 bp
*    45175    45898: contig of 724 bp in length
*    45899 45998: gap of      100 bp
*    45999    46738: contig of 740 bp in length
*    46739 46838: gap of      100 bp
*    46839    47573: contig of 735 bp in length
*    47574 47673: gap of      100 bp
*    47674    48428: contig of 755 bp in length
*    48429 48528: gap of      100 bp
*    48529    49313: contig of 785 bp in length
*    49314 49414: gap of      100 bp
*    49414    50121: contig of 708 bp in length
*    50122 50221: gap of      100 bp
*    50222    50908: contig of 687 bp in length
*    50909 51008: gap of      100 bp
*    51009    51714: contig of 706 bp in length
*    51715 51814: gap of      100 bp
*    51815    52557: contig of 743 bp in length
*    52558 52657: gap of      100 bp
*    52658    53389: contig of 732 bp in length
*    53390 53489: gap of      100 bp
*    53490    54227: contig of 738 bp in length
*    54228 54327: gap of      100 bp
*    54328    55060: contig of 733 bp in length
*    55061 55160: gap of      100 bp
*    55161    55916: contig of 756 bp in length
*    55917 56016: gap of      100 bp
*    56017    56755: contig of 739 bp in length
*    56756 56855: gap of      100 bp
*    56856    57559: contig of 704 bp in length
*    57560 57659: gap of      100 bp
*    57660    58349: contig of 690 bp in length
*    58350 58449: gap of      100 bp
*    58450    59169: contig of 720 bp in length
*    59170 59269: gap of      100 bp
```

```
*    59270    59981: contig of 712 bp in length

alignment scores:
            Quality:    7.00            Length:       7
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY-130_138 x AC010861    ..

 Align seq 1/1  to: AC010861  from: 1  to: 63869

        2 LeuLysAspIleGluThrAla 8
          ||||||||||||||||||||
   61868 CTGAAAGACATAGAAACTGCC 61888

seq name: gb htg5:AC016176

seq documentation block:
LOCUS       AC016176     75254 bp    DNA             HTG       13-JUL-2000
DEFINITION  Homo sapiens clone RP11-2007, LOW PASS SEQUENCE SAMPLING.
ACCESSION   AC016176
VERSION     AC016176.2  GI:9123980
KEYWORDS    HTG; HIGS PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 75254)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-2007
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 75254)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Berezinskiy,I., Bezkubenko,R.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., DeWar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Gallagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,P., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lien,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Testaye,S., Titrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 13, 2000 this sequence version replaced gi:6466618.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
                    -------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www.seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            ---- -------- Project Information
                Center project name: L4959
                Center clone name: 20 0 7
                        .
            * NOTE: This record contains 72 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low pass sequence sampling is useful for
            * identifying clones that may be gene rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *       1     906: contig of 906 bp in length
            *     907    1006: gap of     100 bp
            *    1007    1958: contig of 952 bp in length
            *    1959    2058: gap of     100 bp
            *    2059    2966: contig of 908 bp in length
            *    2967    3066: gap of     100 bp
            *    3067    4025: contig of 959 bp in length
            *    4026    4125: gap of     100 bp
            *    4126    5048: contig of 923 bp in length
            *    5049    5148: gap of     100 bp
            *    5149    6106: contig of 958 bp in length
            *    6107    6206: gap of     100 bp
            *    6207    7157: contig of 951 bp in length
            *    7158    7257: gap of     100 bp
            *    7258    8182: contig of 925 bp in length
            *    8183    8282: gap of     100 bp
            *    8283    9247: contig of 965 bp in length
            *    9248    9347: gap of     100 bp
            *    9348   10277: contig of 930 bp in length
            *   10278   10377: gap of     100 bp
            *   10378   11292: contig of 915 bp in length
            *   11293   11392: gap of     100 bp
            *   11393   12385: contig of 993 bp in length
            *   12386   12485: gap of     100 bp
            *   12486   13417: contig of 932 bp in length
            *   13418   13517: gap of     100 bp
            *   13518   14436: contig of 919 bp in length
            *   14437   14536: gap of     100 bp
            *   14537   15448: contig of 912 bp in length
            *   15449   15548: gap of     100 bp
            *   15549   16527: contig of 979 bp in length
            *   16528   16627: gap of     100 bp
            *   16628   17568: contig of 941 bp in length
            *   17569   17668: gap of     100 bp
            *   17669   18612: contig of 944 bp in length
            *   18613   18712: gap of     100 bp
            *   18713   19655: contig of 943 bp in length
            *   19656   19755: gap of     100 bp
            *   19756   20722: contig of 967 bp in length
            *   20723   20822: gap of     100 bp
            *   20823   21766: contig of 944 bp in length
            *   21767   21866: gap of     100 bp
            *   21867   22818: contig of 952 bp in length
            *   22819   22918: gap of     100 bp
            *   22919   23891: contig of 973 bp in length
            *   23892   23991: gap of     100 bp
            *   23992   24921: contig of 930 bp in length
            *   24922   25021: gap of     100 bp
            *   25022   25957: contig of 936 bp in length
            *   25958   26057: gap of     100 bp
            *   26058   26996: contig of 939 bp in length
            *   26997   27096: gap of     100 bp
            *   27097   28042: contig of 946 bp in length
            *   28043   28142: gap of     100 bp
            *   28143   29092: contig of 950 bp in length
            *   29093   29192: gap of     100 bp
            *   29193   30126: contig of 934 bp in length
            *   30127   30226: gap of     100 bp
            *   30227   31176: contig of 950 bp in length
            *   31177   31276: gap of     100 bp
            *   31277   32236: contig of 960 bp in length
            *   32237   32336: gap of     100 bp
            *   32337   33290: contig of 954 bp in length
            *   33291   33390: gap of     100 bp
            *   33391   34353: contig of 963 bp in length
            *   34354   34453: gap of     100 bp
            *   34454   35371: contig of 918 bp in length
            *   35372   35471: gap of     100 bp
            *   35472   36409: contig of 938 bp in length
            *   36410   36509: gap of     100 bp
            *   36510   37444: contig of 935 bp in length
            *   37445   37544: gap of     100 bp
```

```
*    37545    38441: contig of 897 bp in length
*    38442 38541: gap of       100 bp
*    38542    39481: contig of 940 bp in length
*    39482 39581: gap of       100 bp
*    39582    40525: contig of 944 bp in length
*    40526 40625: gap of       100 bp
*    40626    41602: contig of 977 bp in length
*    41603 41702: gap of       100 bp
*    41703    42634: contig of 932 bp in length
*    42635 42734: gap of       100 bp
*    42735    43665: contig of 931 bp in length
*    43666 43765: gap of       100 bp
*    43766    44744: contig of 979 bp in length
*    44745 44844: gap of       100 bp
*    44845    45805: contig of 961 bp in length
*    45806 45905: gap of       100 bp
*    45906    46900: contig of 995 bp in length
*    46901 47000: gap of       100 bp
*    47001    47938: contig of 938 bp in length
*    47939 48038: gap of       100 bp
*    48039    48971: contig of 933 bp in length
*    48972 49071: gap of       100 bp
*    49072    50019: contig of  948 bp in length
*    50020 50119: gap of       100 bp
*    50120    51066: contig of 947 bp in length
*    51067 51166: gap of       100 bp
*    51167    52097: contig of 931 bp in length
*    52098 52197: gap of       100 bp
*    52198    53186: contig of 989 bp in length
*    53187 53286: gap of       100 bp
*    53287    54206: contig of 920 bp in length
*    54207 54306: gap of       100 bp
*    54307    55236: contig of 930 bp in length
*    55237 55336: gap of       100 bp
*    55337    56324: contig of 988 bp in length
*    56325 56424: gap of       100 bp
*    56425    57423: contig of 999 bp in length
*    57424 57523: gap of       100 bp
*    57524    58456: contig of 933 bp in length
*    58457 58556: gap of       100 bp
*    58557    59493: contig of 937 bp in length
*    59494 59593: gap of       100 bp
*    59594    60536: contig of 943 bp in length
*    60537 60636: gap of       100 bp
*    60637    61577: contig of 941 bp in length
*    61578 61677: gap of       100 bp
*    61678    62623: contig of 946 bp in length
*    62624 62723: gap of       100 bp
*    62724    63699: contig of 976 bp in length
*    63700 63799: gap of       100 bp
*    63800    64751: contig of 952 bp in length
*    64752 64851: gap of       100 bp
*    64852    65807: contig of 956 bp in length
*    65808 65907: gap of       100 bp
*    65908    66806: contig of 899 bp in length
*    66807 66906: gap of       100 bp
*    66907    67860: contig of 954 bp in length
*    67861 67960: gap of       100 bp
*    67961    68910: contig of 950 bp in length
*    68911 69010: gap of       100 bp
*    69011    69963: contig of 953 bp in length
*    69964 70063: gap of       100 bp
*    70064    71006: contig of 943 bp in length
*    71007 71106: gap of       100 bp
*    71107    72081: contig of 975 bp in length
*    72082 72181: gap of       100 bp
*    72182    73147: contig of 966 bp in length
*    73148 73247: gap of       100 bp
*    73248    74168: contig of 921 bp in length
*    74169 74268: gap of       100 bp
*    74269    75254: contig of 986 bp in length.
FEATURES             Location/Qualifiers
     source          1..75254
```

```
alignment_scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_130_138 x AC016176/rev  ..

 Align seq 1/1  to reverse of: AC016176  from: 1  to: 75254

       2 LeuTyrAspIleGluThrAla 8
         |:|: |||||||||||||:||
   47881 CTAAACGATATAGAAACGGT 47861
```

```
OM of: US-09-126-945B-2_COPY_130_138.fas* N_Geneseq_0601.* ***** pts

Date: Sep 14, 2001  9:48 PM

About: Results were produced by the [illegible]
       Copyright (c) 1993-2000 Compugen [illegible]

Command line parameters:
  -MODEL frame+_p2n.model -DEV-xlp
  -Q "us-09-126-945b-2 copy 130 138 olip2n.rng" -SEQ pct "/cgn2_1/ptodata/1/pubpct/US09_126_945B_2_COPY_130_138_OLIP2N.RNG" query 1 1087
  -DB N_Geneseq_0601 -QFMT=fastap -SUFFIX olip2n.rng -GAPOP 4.500
  -GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT 0.000
  -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT 60.000
  -FGAPOP 6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT 60.000
  -DELOP 6.000 -DELEXT=7.000 -START=1 -MATRIX align
  -TRANS human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE quality
  -THR_MIN=1 -ALIGN 15 -MODE LOCAL -OUTFMT pts  NORM=ext  MINLEN 0
  -MAXLEN 2000000000  USER=US09126945_2COPY_130_138  NCPU 6  JOBID 4
  -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-126-945B-2_COPY_130_138
Query length: 9
Database: N_Geneseq_0601.*
Database sequences: 730101
Database length: 314950809
Search time (sec): 680.150000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000


score_list:
Sequence            Strd Orig    ZScore   EScore Len  ! Documentation  ..
/SIDS1/qcqdata/geneseq/geneseqn/NA2000 DAT:AAF21828  -     6.00  128.19   36.04     18  ! A
/SIDS1/qcqdata/geneseq/geneseqn/NA1995 DAT:AAI08218  -     5.00  127.14   41.42     21  ! A
/SIDS1/qcqdata/geneseq/geneseqn/NA2000 DAT:AAA07229  -     5.00  125.98   51.67     27  ! B
/SIDS1/qcqdata/geneseq/geneseqn/NA1995 DAT:AAI26090  -     5.00  116.97   152.19    91  !
/SIDS1/qcqdata/geneseq/geneseqn/NA2000 DAT:AAC51119  -     5.00  111.43   224.41    136
/SIDS1/qcqdata/geneseq/geneseqn/NA2000 DAT:AAC52079  -     5.00  111.43   292.54    179
/SIDS1/qcqdata/geneseq/geneseqn/NA2000 DAT:AAB09277  -     5.00  111.08   324.81    213
/SIDS1/qcqdata/geneseq/geneseqn/NA1997 DAT:AAT68052  -     5.00  110.79   345.93    222
/SIDS1/qcqdata/geneseq/geneseqn/NA2000 DAT:AAC28909  -     5.00  111.43   445       229

seq_type:  "AA" "/SIDS1/qcqdata/geneseq/geneseqn/NA2000.DAT:AAF21828

seq_documentation_block:
ID   AAF21828 standard; DNA; 1087 BP.
XX
AC   AAF21828;
XX
DT   27-MAR-2001  (first entry)
XX
DE   Human breast and ovarian cancer associated antigen gene SEQ ID 215.
XX
KW   Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
KW   nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW   antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
KW   antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
KW   Addison's disease; allergy; autoimmune haemolytic anaemia;
KW   autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW   multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW   cardiovascular disorder; wound healing; neurological disease; ds.
XX
OS   Homo sapiens.
XX
PN   WO200055177-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05881.
XX
PR   12-MAR-1999;  99US-0124270.
XX
PA   (HEMA ) HUMAN GENOME SCI INC.
XX
PI   [illegible]
XX
DR   WPI; [illegible]
DR   P-PSDB; AAB58925.
XX
PT   New human breast and ovarian cancer associated gene sequences and the
PT   polypeptides encoded by these genes, useful in the prevention,
PT   treatment and diagnosis of cancer, immune disorders, cardiovascular
PT   disorders and neurological diseases -
XX
PS   Claim 1; Page 642-643; 1299pp; English.
XX
CC   Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
CC   proteins AAB59711 - AAB59128. The DNA and protein sequences are
CC   associated with breast and ovarian cancer. Included in the invention are
CC   sequences AAF21997 - AAF22040 and AAB59129 which are used in the
CC   isolation and characterisation of the DNA and protein sequences of the
CC   invention. The breast and ovarian cancer associated DNA, protein, agonist
CC   [illegible] sequences exhibit cytostatic; immunosuppressive;
CC   nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
CC   antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
CC   antibacterial; antifungal; antiparasitic; cardiant activity. The
CC   polynucleotide and protein sequences are used in the diagnosis, [illegible] cancer,
CC   particularly breast and ovarian cancer. The nucleic acid sequences,
CC   proteins, agonists and agonists may also be used in the diagnosis,
CC   prevention and treatment of immune disorders e.g. Addison's disease,
CC   allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC   diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC   arthritis, [illegible] cardiovascular disorders such as
CC   myocardial ischaemias, wound healing, neurological diseases such as
CC   cerebral ataxia and epilepsy, and infectious diseases.
XX
SQ   Sequence 1087 BP; 215 A; 447 C; 339 G; 182 T; 4 other;
```

```
alignment scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2 COPY_130_138 x AAF21828    .

 Align_seq_1/1  to: AAF21828  from: 1  to: 1087

        1 ValTleuLysAspIleGluThrAlaCys 9
          |||||||||||||||||||||||||||
      264 GTGCTCAAGGACATCGAGACGGCCTGC 290

Seq name: /cgn2_6/ptodata/1/pubseq/seqdata/pc/SA2500.DAT:AAZ50691

Seq documentation block:
ID   AAZ50691 standard; cDNA; 1894 BP.
XX
AC   AAZ50691;
XX
DT   24-MAY-2000  (first entry)
XX
DE   Human prostate derived Ets factor cDNA.
XX
KW   Prostate derived Ets factor; PDEF; human; chromosome 6p21.3; cancer;
KW   loss of heterozygosity; chromosomal translocation; linkage analysis;
KW   cytostatic; cardiant; immunosuppressive; cerebroprotective; fungicide;
KW   antibacterial; vulnerary; neuroprotective; antiparkinsonian; nootropic;
KW   anabolic; antiinflammatory; anorectic; hybridisation probe; forensic;
KW   tumour marker, diagnosis, treatment, prostate cancer, blood coagulation;
KW   autoimmune disorder; haematopoietic; immune/nervous system; stroke;
KW   neoplasm; microbial infection; tissue regeneration; heart attack;
KW   scarring; food additive; preservative; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             416..1423
FT                   /*tag= a
FT                   /product= "Human prostate derived Ets factor"
FT                   /note= "Expressed in prostate epithelium"
XX
PN   WO200006589-A1.
XX
PD   10-FEB-2000.
XX
PF   02-AUG-1999;   99WO-US17470.
XX
PR   31-JUL-1998;   98US-0126945.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI   Libermann TA, Oettgen JP, Kunsch CA, Endress GA, Rosen CA;
XX
DR   WPI; 2000-195255/17.
DR   P-PSDB; AAY45004.
XX
PT   Novel prostate derived polypeptide, polynucleotide useful for
PT   diagnosis, prevention and treatment of prostate cancer, autoimmune
PT   disorders, microbial infections and also as food additive or
PT   preservative
XX
PS   Claim 4; Fig 1; 132pp; English.
XX
CC   The present cDNA sequence encodes a human prostate derived Ets factor
CC   (PDEF), isolated from human prostate carcinoma cDNA library
CC   (ATCC No.203072). PDEF gene expression is highest in tissues involved in
CC   androgen sensitivity, with enriched expression in prostate epithelium
CC   cells. PDEF gene is mapped to the human chromosome 6p21.3
CC   region that is associated with loss of heterozygosity and chromosomal
CC   translocations in various human cancers. PDEF has cytostatic, cardiant,
CC   immunosuppressive, cerebroprotective, fungicide, antibacterial,
CC   vulnerary, neuroprotective, antiparkinsonian, nootropic, anabolic,
CC   antiinflammatory and anorectic activity. PDEF polynucleotides are useful
CC   in linkage analysis as markers, as hybridisation probes for differential
CC   identification of the tissues or cell types and as polymorphic markers
CC   for forensic purposes. PDEF is useful as prostate-specific tumour marker
CC   for the diagnosis and treatment of prostate cancer. PDEF sequences are
CC   useful for treating autoimmune disorders, haematopoietic, blood
CC   coagulation, immune and nervous system disorders, hyperproliferative
CC   disorders like, neoplasms and microbial infections, heart attacks,
CC   stroke, scarring and for tissue regeneration. They are also useful as
CC   food additives or preservatives.
XX
SQ   Sequence 1894 BP; 368 A; 654 C; 571 G; 302 T; 0 other;


alignment scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2 COPY_130_138 x AAZ50691    ..

 Align_seq_1/1  to: AAZ50691  from: 1  to: 1894

        1 ValTleuLysAspIleGluThrAlaCys 9
          |||||||||||||||||||||||||||
      803 GTGCTCAAGGACATCGAGACGGCCTGC 829

Seq name: /cgn2_6/ptodata/1/pubseq/seqdata/pc/NA2001.DAT:AAC83261

Seq documentation block:
ID   AAC83261 standard; DNA; 1894 BP.
XX
AC   AAC83261;
XX
DT   16-MAR-2001  (first entry)
XX
DE   Gland specific Ets transcription factor (GSEF) nucleotide sequence.
XX
KW   Transcription factor; gland specific Ets transcription factor; GSEF;
KW   metastatic potential; cancer; tumour; metastasis; breast; prostate;
KW   leukaemia; lymphoma; sarcoma; melanoma; chromosome 6p21.1-6p21.3; ds.
XX
OS   Homo sapiens.
XX
PN   WO200070092-A1.
XX
PD   23-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US13173.
XX
PR   14-MAY-1999;   99US-0134112.
XX
PA   (CHIR-) CHIRON CORP.
XX
PI   Kaufmann I, Xiu H, Harrowe G;
XX
DR   WPI; 2001-041019/05.
DR   P-PSDB; AAB49628.
XX
PT   Detecting metastatic and potential metastatic cancerous cells, useful
PT   for diagnosing, prognosing, grading and staging of cancers by detecting
PT   gland specific Ets transcription factor gene product in a biological
PT   sample from a cell
XX
PS   Disclosure; Page 81-83; 95pp; English.
XX
CC   This invention relates to a method for the detection and determination of
CC   the metastatic potential of a cell. The method comprises detecting a
CC   gland specific Ets transcription factor (GSEF) gene product in a test
```

```
CC   sample. Detection of a GSEF gene product in the test sample in amount
CC   lower than that in a normal cell, is indicative of a cell with high
CC   metastatic potential. The method is useful for determining the metastatic
CC   potential of a cell, for the diagnosis and prognosis of cancer as well as
CC   grading and staging of cancers by detecting GSEF expression in a
CC   biological test sample. The method may also be used to monitor patients
CC   having a predisposition to develop a particular cancer. GSEF polypeptides
CC   are useful for producing antibodies, in cancer diagnosis, prognosis,
CC   grading, staging and management of breast and prostate tumours, and in
CC   detecting polymorphisms in the sequence. GSEF genes and proteins are also
CC   useful in gene therapy. GSEF gene product expression levels can be used
CC   in conjunction with any tissue in which an alteration in GSEF gene
CC   product expression levels is associated with development of a
CC   cancer associated phenotype. Cancers, which can be monitored include
CC   cancers of the prostate, cervix, lung and colon, melanomas, colorectal
CC   adenocarcinomas, Wilm's tumour, retinoblastoma, sarcomas, myosarcomas,
CC   lung carcinomas, leukaemia, and lymphomas. The GSEF gene is located on
CC   human chromosome 6, specifically at 6p21.1-6p21.3. The present sequence
CC   represents the DNA encoding GSEF
XX
SQ   Sequence 1894 BP; 368 A; 653 C; 571 G; 302 T; 0 other;


alignment_scores;
              Quality:    9.00              Length:       9
                Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_130_138 x AAC83261    ..

 Align seq 1/1  to: AAC83261  from: 1  to: 1894

        1 ValLeuLysAspIleGluThrAlaCys 9
          ||||||||||||||||||||||||||
      803 GTGCTCAAGGACATCGAGACGGCCTGC 829

Seq_name: /cit61/6-pdata/pep-seq-protein-pg/NA.DAT:AAC83266

Seq_documentation_block:
ID   AAC83266 standard; DNA; 3317 BP.
XX
AC   AAC83266;
XX
DT   16-MAR-2001  (first entry)
XX
DE   Gland-specific Ets transcription factor (GSEF) cDNA sequence.
XX
KW   Transcription factor; gland-specific Ets transcription factor; GSEF;
KW   metastatic potential; cancer; tumour; metastasis; breast; prostate;
KW   leukaemia; lymphoma; sarcoma; melanoma; chromosome 6p21.1-6p21.3; ss.
XX
OS   Homo sapiens.
XX
PN   WO200070092-A1.
XX
PD   23-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US13174.
XX
PR   14-MAY-1999;   99US-0134112.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Kaufmann J;  Xin H;  Harrowe G;
XX
DR   WPI; 2001-041819/05.
DR   P-PSDB; AAB49633.
XX
PT   Detecting metastatic and potential metastatic cancerous cells, useful
PT   for diagnosing, prognosing, grading and staging of cancers by detecting
PT   gland-specific Ets transcription factor gene product in a biological
PT   sample from a cell -
XX
PS   Disclosure; Fig 2; 95pp; English.
XX
CC   This invention relates to a method for the detection and determination of
CC   the metastatic potential of a cell. The method comprises detecting a
CC   gland-specific Ets transcription factor (GSEF) gene product in a test
CC   sample. Detection of a GSEF gene product in the test sample in amount
CC   lower than that in a normal cell, is indicative of a cell with high
CC   metastatic potential. The method is useful for determining the metastatic
CC   potential of a cell, for the diagnosis and prognosis of cancer as well as
CC   grading and staging of cancers by detecting GSEF expression in a
CC   biological test sample. The method may also be used to monitor patients
CC   having a predisposition to develop a particular cancer. GSEF polypeptides
CC   are useful for producing antibodies, in cancer diagnosis, prognosis,
CC   grading, staging and management of breast and prostate tumours, and in
CC   detecting polymorphisms in the sequence. GSEF genes and proteins are also
CC   useful in gene therapy. GSEF gene product expression levels can be used
CC   in conjunction with any tissue in which an alteration in GSEF gene
CC   product expression levels is associated with development of a
CC   cancer-associated phenotype. Cancers, which can be monitored include
CC   cancers of the prostate, cervix, lung and colon, melanomas, colorectal
CC   adenocarcinomas, Wilm's tumour, retinoblastoma, sarcomas, myosarcomas,
CC   lung carcinomas, leukaemia, and lymphomas. The GSEF gene is located on
CC   human chromosome 6, specifically at 6p21.1-6p21.3. The present sequence
CC   represents the cDNA encoding GSEF.
XX
SQ   Sequence 3317 BP; 710 A; 1026 C; 970 G; 611 T; 0 other;


alignment_scores;
              Quality:    9.00              Length:       9
                Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_130_138 x AAC83266    .

 Align seq 1/1  to: AAC83266  from: 1  to: 3317

        1 ValLeuLysAspIleGluThrAlaCys 9
          ||||||||||||||||||||||||||
     2202 GTGCTCAAGGACATCGAGACGGCCTGC 2228

Seq_name: /cit61/6-pdata/pep-seq-protein-pg/NA1988.DAT:AAN81720

Seq_documentation_block:
ID   AAN81720 standard; DNA; 885 BP.
XX
AC   AAN81720;
XX
DT   22-OCT-1990  (first entry)
XX
DE   Sequence encoding immunoreactive human B-lymphotropic viral protein.
XX
KW   Immunoreactive; human B lymphotropic virus; ss.
XX
OS   synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..885
FT                   /*tag= a
XX
PN   WO8809814-A.
XX
PD   15-DEC-1988.
XX
PF   27-MAY-1988;    88WO-US01802.
XX
PR   01-JUN-1987;    87US-0056963.
XX
PA   (BAYU ) BAYLOR UNIV COLLEGE.
PA   (USSR ) US DEPT OF HEALTH AND HUMAN SERVICES.
XX
```

PT Chang NT; Chang TW; Fung MS C; Fung M C; Gallo RC; Wong-Staal F;
XX
DR WPI; 1988-368628/51.
DR P-PSDB; AAP81328.
XX
PT New immunoreactive protein of human B lymphotropic virus and encoding
PT DNA sequences, useful as immunogens and for detecting specific antibodies
PT in blood
XX
PS Disclosure; 5 ; English.
XX
CC This is part of the human B lymphotropic virus (HBLV) genome and encodes
CC an immunoreactive protein which can be used to detect antibodies against
CC HBLV in biological fluids (esp. for screening human serum or plasma).
XX
SQ Sequence 885 BP; 465 A; 214 C; 124 G; 182 T; 0 other;

alignment scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2 (COPY 130-138 x AAN81720 ..

Align Seq 1/1 to: AAN81720 from: 1 to: 885

1 ValGluLysAspIleGluThr 7
|||||||||||||||||||||
286 GTGGTTAAAAGACATAGAAACG 306

seq_name; /cgi-bin/pdbs/geneseq/geneseqp-NA2000.DAT:AAC41025

seq documentation block:
ID AAC41025 standard; cDNA; 96 BP.
XX
AC AAC41025;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 45100.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J; Duclert A; Giordano J;
XX
DR WPI; 2000-500381/45.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures
XX
PS Claim 1; SEQ ID 45100; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 40 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
SQ Sequence 96 BP; 34 A; 11 C; 14 G; 37 T; 0 other;

alignment scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2_COPY 130-138 x AAC41025 ..

Align Seq 1/1 to: AAC41025 from: 1 to: 96

8 LysAspIleGluThrAla 8
|||||||||||||||||||
55 AAACATATAGAAACTGCA 72

seq_name; /cgi-bin/pdbs/geneseq/geneseqp-NA2000.DAT:AAC05069

seq_documentation_block:
ID AAC05069 standard; cDNA; 215 BP.
XX
AC AAC05069;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 9144.
XX
KW Human, 5' EST, expressed sequence tag, secreted protein, cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J; Duclert A; Giordano J;
XX
DR WPI; 2000-500381/45.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures
XX
PS Claim 1; SEQ ID 9144; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 40 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be

CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
SQ Sequence 215 BP; 44 A; 37 C; 41 G; 104 T; 0 other;

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_130_138 x AAC05069/rev .

Align seq 1/1 to reverse of: AAC05069 from: 1 to: 215

2 ProLysAspIleGluThr 7
|||||||||||||||||||
83 CTGAAGGACATAGAGACA 66

Seq name: /SIDS2/gcgdata/geneseqp/geneseqp_NA.*, 1 of AAC12179

seq_documentation_block:
ID AAZ42413 standard; cDNA; 218 BP.
XX
AC AAZ42413;
XX
DT 01-FEB-2000 (first entry)
XX
DE Human 5' EST isolated from a cDNA library SEQ ID No:172.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
KW gene therapy; chromosome mapping; upstream regulatory sequence;
KW forensic; location; development; protein synthesis; stability;
KW regulation; identification; ss.
XX
OS Homo sapiens.
XX
PN WO9953051-A2.
XX
PD 21-OCT-1999.
XX
PF 09-APR-1999; 99WO-IB00712.
XX
PR 09-APR-1998; 98US-0057719.
PR 28-APR-1998; 98US-0069047.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-038446/03.
DR P-PSDB; AAY64799.
XX
PT Novel secreted protein 5' expressed sequence tag sequences used in
PT diagnostic, forensic, gene therapy, and chromosome mapping procedures
XX
PS Claim 1; Page 267; 837pp; English.
XX
CC AAZ42265 to AAZ43075 represent novel 5' expressed sequence tag (EST)
CC sequences, corresponding to human secreted proteins. AAY64951 to
CC AAY65438 represent the EST-related proteins corresponding to AAZ42265 to
CC AAZ43052. The 5' ESTs can be used for producing secreted human gene
CC products. They can be used to identify and isolate 5' untranslated
CC regions (UTRs) and upstream regulatory regions which control the
CC location, development stage, rate, and quantity of protein synthesis, as
CC well as stability of mRNA. The ESTs are also useful as probes for
CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can
CC also be used in forensic procedures to identify individuals, or in
CC diagnostic procedures to identify individuals having genetic diseases
CC resulting from abnormal gene expression. The products may also be used in
CC gene therapy protocols. The nucleic acids encoding signal peptides can be
CC used for directing extracellular secretion of a polypeptide or the
CC insertion of a polypeptide into a membrane, or importing a polypeptide
CC into a cell. The proteins encoded by the EST sequences may be useful in
CC treating a variety of human conditions. Secreted proteins have
CC therapeutic value, and the identification of new secreted proteins is
CC valuable. AAZ42249 to AAZ42264 and AAY64644 to AAY64650 represent
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 218 BP; 44 A; 37 C; 44 G; 103 T; 0 other;

alignment scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2_COPY_130_138 x AAZ42413/rev ..

Align seq 1/1 to reverse of: AAZ42413 from: 1 to: 218

2 ProLysAspIleGluThr 7
|||||||||||||||||||
83 CTGAAGGACATAGAGACA 66

Seq name: /SIDS2/gcgdata/geneseqp/geneseqp_NA*/geneseqp_NA*.seq of AAZ33728

seq_documentation_block:
ID AAZ33728 standard; cDNA; 573 BP.
XX
AC AAZ33728;
XX
DT 09-DEC-1999 (first entry)
XX
DE Tobacco plant resistance associated cDNA fragment 54.
XX
KW Tobacco; plant; resistance; tobacco mosaic virus; TMV; infection;
KW protection; plant protection agent; phytopathogenic fungi; nematode;
KW insect pest; pathogen resistance; transgenic plant; diagnosis; ss.
XX
OS Nicotiana tabacum.
XX
PN DE19813048 A1.
XX
PD 30-SEP-1999.
XX
PF 25-MAR-1998; 98DE 1013048.
XX
PR 25-MAR-1998; 98DE 1013048.
XX
PA (FARB) BAYER AG.
XX
PI Schreier P, Gartemann C, Landen G, Kiedrowski S;
XX
DR WPI; 1999-552163/47.
XX
PT Nucleic acids useful for identifying compounds capable of inducing
PT pathogen resistance in plants -
XX
PS Claim 1a; 40; 130pp; German.
XX
CC This invention describes novel cDNA molecules corresponding to Tobacco
CC (Nicotiana tabacum) genes that are expressed in response to tobacco
CC mosaic virus (TMV) infection. Regulatory regions that provide specific
CC expression of the nucleic acids in connection with induced resistance
CC can be used to identify compounds useful for plant protection by
CC transforming a cell with a recombinant DNA molecule containing a
CC selectable and/or detectable marker linked to the regulatory region,
CC culturing the cell in the presence of a test compound or sample, and
CC identifying a compound or sample that suppresses or activates and/or
CC enhances expression of the marker. Host cells containing the nucleic
CC acid, or polypeptides encoded by the nucleic acids, can be used to

CC identify compounds that inhibit or activate a polypeptide involved in
CC induced resistance, by contacting the cell or polypeptide with one or
CC more test compounds and identifying compound(s) that specifically
CC modulate the activity of the polypeptide. Plant protection agents can be
CC prepared by formulating the identified compounds in a form suitable for
CC their uptake and optionally metabolization in organisms, especially
CC plants, phytopathogenic fungi, nematodes or insect pests. The nucleic
CC acids or regulatory regions, or corresponding vectors, or the
CC polypeptides encoded by the nucleic acids, or compounds identified as
CC above, can be used to improve the pathogen resistance of plants or to
CC produce transgenic plants. The nucleic acids or regulatory regions, or
CC vectors, cells, plants or plant tissues containing them, or the
CC polypeptides and antibodies to the polypeptides, can be used to screen
CC for new plant protection agents or genes involved in induction of
CC resistance in plants. Vectors containing the nucleic acids, or the
CC regulatory regions or recombinant DNA molecule or vectors containing
CC them, or the compounds identified as above, or the polypeptides or
CC antibodies, can be used in diagnostic compositions. AAZ33676 Z33868
CC represent fragments of tobacco cDNA which are associated with plant
CC resistance.
XX
SQ Sequence 473 BP; 122 A; 60 C; 78 G; 113 T; 0 other;

alignment scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2-COPY-130_138 x AAZ33728 ..

Align seq 1/1 to: AAZ33728 from: 1 to: 473

3 LysAspIleGluThrAla 8
||||||||||||||||||
103 AAAGATATTGAAACTGCG 120

seq name: /cgi-bin/sp-data/gene_seq?genesep:pNA2000:DAT:AAF09959

seq documentation block:
ID AAF09959 standard; cDNA; 509 BP.
XX
AC AAF09959;
XX
DT 13-MAR-2001 (first entry)
XX
DE Fusarium venenatum EST SEQ ID NO:4482.
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Fusarium venenatum.
XX
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0274623.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
PA (NOVO) NOVO NORDISK AS.
XX
PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olson PB;
XX
DR WPI; 2000-594572/56.
XX
PT Monitoring differential expression of genes in filamentous fungal cells

PT uses fluorescence labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
XX
PS Claim 86; Page 1293; 816pp; English.
XX
CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (ESTs). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Mor -ring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.
XX
SQ Sequence 509 BP; 119 A; 117 C; 116 G; 150 T; 7 other;

alignment scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2-COPY-130_138 x AAF09959/rev ..

Align seq 1/1 to reverse of: AAF09959 from: 1 to: 509

2 LeuLysAspIleGluThr 7
||||||||||||||||||
489 CTTAAAGATATAGAAACA 472

seq name: /cgi-bin/sp-data/gene_seq?genesep:pNA2000:DAT:AAF11962

seq documentation block:
ID AAF11962 standard; cDNA; 736 BP.
XX
AC AAF11962;
XX
DT 13-MAR-2001 (first entry)
XX
DE Aspergillus oryzae EST SEQ ID NO:4485.
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Aspergillus oryzae.
XX
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0274623.

```
XX
PA   (NOVO ) NOVO NORDISK BIOTECH INC.
PA   (NOVO ) NOVO NORDISK AS.
XX
PI   Berka PM, Rey MW, Shuster JR, Kroppinon S, Clausen IG, Olson PB;
XX
DR   WPI; 2000-594572/56.
XX
PT   Monitoring differential expression of genes in filamentous fungal cells
PT   uses fluorescence-labeled nucleic acids isolated from the cells and a
PT   substrate of expressed sequence tags.
XX
PS   Claim 88; Page 1914; 316lpp; English.
XX
CC   The present invention describes a method for monitoring differential
CC   expression of genes in a first filamentous fungal (Ff) cell relative to
CC   expression of the same genes in one or more second filamentous fungal
CC   cells. The method uses fluorescence-labeled nucleic acids isolated from
CC   the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC   are used in the methods for monitoring differential expression of genes
CC   in a first filamentous fungal (FF) cell relative to expression of the
CC   same genes in one or more second filamentous fungal cells. Monitoring
CC   the global expression of genes from FF cells allows the production
CC   potential of the microorganisms to be improved. New genes may be
CC   discovered, possible functions of unknown open reading frames can be
CC   identified and gene copy number variation and stability can be
CC   monitored. The expression of genes can be used to study how FF cells
CC   adapt to changes in culture conditions, environmental stress, spore
CC   morphogenesis, recombination, metabolic or catabolic pathway
CC   engineering. Using ESTs provides several advantages over genomic or
CC   random cDNA clones including elimination of redundancy as one spot on an
CC   array equals one gene or open reading frame, and organisation of the
CC   microarrays based on function of the gene products to facilitate
CC   analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC   Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC   niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC   AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC   all specifically claimed in the present invention.
XX
SQ   Sequence 736 BP; 160 A; 163 C; 198 G; 214 T; 1 other;


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_130_138 x AAF11962  ..

 Align seq 1/1  to: AAF11962  from: 1  to: 736

       1 ValLeuLysAspIleGln 6
         ||||||||||||||||||
     423 GTCCTTAAGGATATTGAG 440

seq name: /SIDSI/Gen3ita/gen3seq/gen3seq_p_NS. ** IAC AAC52278

seq_documentation_block:
ID   AAC52278 standard; DNA; 1026 BP.
XX
AC   AAC52278;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 70827.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   24-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
```

PR 01 JUL 1999; 99US-0142154.
PR 02 JUL 1999; 99US-0142055.
PR 06 JUL 1999; 99US-0142390.
PR 08 JUL 1999; 99US-0142803.
PR 09 JUL 1999; 99US-0142920.
PR 12 JUL 1999; 99US-0142977.
PR 13 JUL 1999; 99US-0143542.
PR 14 JUL 1999; 99US-0143624.
PR 15 JUL 1999; 99US-0144005.
PR 16 JUL 1999; 99US-0144085.
PR 16 JUL 1999; 99US-0144086.
PR 19 JUL 1999; 99US-0144325.
PR 19 JUL 1999; 99US-0144331.
PR 19 JUL 1999; 99US-0144332.
PR 19 JUL 1999; 99US-0144333.
PR 19 JUL 1999; 99US-0144334.
PR 19 JUL 1999; 99US-0144335.
PR 20 JUL 1999; 99US-0144352.
PR 20 JUL 1999; 99US-0144632.
PR 20 JUL 1999; 99US-0144884.
PR 21 JUL 1999; 99US-0144814.
PR 21 JUL 1999; 99US-0145086.
PR 21 JUL 1999; 99US-0145088.
PR 22 JUL 1999; 99US-0145085.
PR 22 JUL 1999; 99US-0145087.
PR 22 JUL 1999; 99US-0145089.
PR 22 JUL 1999; 99US-0145192.
PR 23 JUL 1999; 99US-0145145.
PR 23 JUL 1999; 99US-0145218.
PR 23 JUL 1999; 99US-0145224.
PR 26 JUL 1999; 99US-0145276.
PR 27 JUL 1999; 99US-0145913.
PR 27 JUL 1999; 99US-0145918.
PR 27 JUL 1999; 99US-0145919.
PR 28 JUL 1999; 99US-0145951.
PR 02 AUG 1999; 99US-0146386.
PR 02 AUG 1999; 99US-0146388.
PR 02 AUG 1999; 99US-0146389.
PR 03 AUG 1999; 99US-0147038.
PR 04 AUG 1999; 99US-0147204.
PR 04 AUG 1999; 99US-0147302.
PR 05 AUG 1999; 99US-0147192.
PR 05 AUG 1999; 99US-0147260.
PR 06 AUG 1999; 99US-0147303.
PR 06 AUG 1999; 99US-0147416.
PR 09 AUG 1999; 99US-0147493.
PR 09 AUG 1999; 99US-0147945.
PR 10 AUG 1999; 99US-0148171
PR 11 AUG 1999; 99US-0148319.
PR 12 AUG 1999; 99US-0148341.
PR 13 AUG 1999; 99US-0148565.
PR 13 AUG 1999; 99US-0148684.
PR 16 AUG 1999; 99US-0149368.
PR 17 AUG 1999; 99US-0149175.
PR 18 AUG 1999; 99US-0149426.
PR 20 AUG 1999; 99US-0149722.
PR 20 AUG 1999; 99US-0149724.
PR 20 AUG 1999; 99US-0149929.
PR 23 AUG 1999; 99US-0149902.
PR 23 AUG 1999; 99US-0149930.
PR 25 AUG 1999; 99US-0150566.
PR 26 AUG 1999; 99US-0150884.
PR 27 AUG 1999; 99US-0151065.
PR 27 AUG 1999; 99US-0151066.
PR 27 AUG 1999; 99US-0151080.
PR 30 AUG 1999; 99US-0151303.
PR 31 AUG 1999; 99US-0151438.
PR 01 SEP 1999; 99US-0151930.
PR 07 SEP 1999; 99US-0152364.
PR 10 SEP 1999; 99US-0153070.
PR 13 SEP 1999; 99US-0153758.
PR 15 SEP 1999; 99US-0154018.
PR 16 SEP 1999; 99US-0154039.
PR 20 SEP 1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157754.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945b-2_copy_130_138 x AAC52278/rev ..

Align set 1/1 to reverse set: AAC52278 from: 1 to: 1026

1 ValLeuLysAspGlnGlu 6
|||| ||||||||||||
444 GTACTCAAAGACATAGAG 427

seq_con [illegible] pCNA2000:DA1:AAC46174

Seq_documentation block:
ID AAC46174 standard; DNA; 1074 BP.
XX
AC AAC46174;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 49168.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.

XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301449.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 24-MAR-1999; 99US-0125788
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 21-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138647.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142804.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145914.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147494.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151304.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152864.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0154758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.

```
PR   22 SEP 1999;   99US-0155149.
PR   23 SEP 1999;   99US-0155486.
PR   24 SEP 1999;   99US-0155659.
PR   28 SEP 1999;   99US-0156458.
PR   29 SEP 1999;   99US-0156596.
PR   04 OCT 1999;   99US-0157117.
PR   05 OCT 1999;   99US-0157754.
PR   06 OCT 1999;   99US-0157865.
PR   07 OCT 1999;   99US-0158029.
PR   08 OCT 1999;   99US-0158242.
PR   12 OCT 1999;   99US-0158369.
PR   13 OCT 1999;   99US-0159293.
PR   13 OCT 1999;   99US-0159294.
PR   13 OCT 1999;   99US-0159295.
PR   14 OCT 1999;   99US-0159329.
PR   14 OCT 1999;   99US-0159330.
PR   14 OCT 1999;   99US-0159341.
PR   14 OCT 1999;   99US-0159647.
PR   14 OCT 1999;   99US-0159648.
PR   18 OCT 1999;   99US-0159584.
PR   21 OCT 1999;   99US-0160741.
PR   21 OCT 1999;   99US-0160767.
PR   21 OCT 1999;   99US-0160768.
PR   21 OCT 1999;   99US-0160770.
PR   21 OCT 1999;   99US-0160814.
PR   21 OCT 1999;   99US-0160815.
PR   22 OCT 1999;   99US-0160980.
PR   22 OCT 1999;   99US-0160981.
PR   22 OCT 1999;   99US-0160989.
PR   25 OCT 1999;   99US-0161404.
PR   25 OCT 1999;   99US-0161405.
PR   25 OCT 1999;   99US-0161406.
PR   26 OCT 1999;   99US-0161359.
PR   26 OCT 1999;   99US-0161360.
PR   26 OCT 1999;   99US-0161361.
PR   28 OCT 1999;   99US-0161920.
PR   28 OCT 1999;   99US-0161992.
PR   28 OCT 1999;   99US-0161993.
PR   29 OCT 1999;   99US-0162142.

alignment scores:
            Quality:    6.00              Length:      6
              Ratio:   1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_130_138 x AAC46174    ..

 Align seq 1/1  to: AAC46174  from: 1  to: 1074

        4 LysAspIleGluThrAla 8
          ||||||||||||||||||
      548 AAACACATAGAAACAGCT 565

seq name: /cgn2_6/prodata/2/genseq/seqdata/pCNA... 1/1/AAC46174

Seq documentation block:
ID   AAA44456 standard; cDNA; 1077 BP.
XX
AC   AAA44456;
XX
DT   21-AUG-2000  (first entry)
XX
DE   Human secreted expressed sequence tag SEQ ID NO:1041.
XX
KW   Human; mouse; chicken; rat; secreted expressed sequence tag; sEST;
KW   expressed sequence tag; EST; probe; chemotactic; proliferative;
KW   immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
KW   thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
KW   antiviral; antidiabetic; antiasthmatic; vulnerary; antiparkinsonian;
KW   antiulcer; osteopathic; neuroprotective; nootropic; antipsoriatic;
KW   cerebroprotective; anticonvulsant; antidepressant; gene therapy;
KW   vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
KW   insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
KW   lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
KW   central nervous system disorder; Alzheimer's disease; stroke;
KW   Parkinson's disease; Huntington's disease; coagulation disorder;
KW   haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
KW   tumour; infection; depression; psoriasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200021991-A1.
XX
PD   20-APR-2000.
XX
PF   15-OCT-1999;   99WO-US24206.
XX
PR   15-OCT-1998;   98US-0104436.
XX
PA   (GEMY ) GENELTS INST INC.
XX
PI   Jacobs K,  McCoy JM,  LaValle ER,  Collins Kacie LA,  Evans C;
PI   Merberg D,  Treacy M,  Bowman MR;
XX
DR   WPI; 2000-317938/27.
XX
PT   Isolated polynucleotides, and encoded proteins, comprising secreted
PT   expressed sequence tags (SESTs), useful for treating various disorders
PT   such as autoimmune, infectious, and central nervous system disorders.
XX
PS   Claim 1; Page 467; 803pp; English.
XX
CC   AAA44426 to AAA45925 represent specifically claimed secreted expressed
CC   sequence tags (sESts), isolated from human, mouse, chicken and rat
CC   tissue sources. The sESts can have a range of activities depending on
CC   the tissues they were isolated from. The activities include:
CC   chemotactic; proliferative; immunomodulatory; haematopoietic;
CC   chemokinetic, analgesic, haemostatic; thrombolytic; antiinflammatory;
CC   cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
CC   antiasthmatic; vulnerary; antiulcer; osteopathic; neuroprotective;
CC   nootropic, antiparkinsonian, antipsoriatic, cerebroprotective;
CC   anticonvulsant; and antidepressant. The sESTs can be used for gene
CC   therapy and in vaccines. The sESTs are useful as probes for the
CC   identification and isolation of full length cDNAs and genomic DNA
CC   molecules which correspond to the sESTs. Proteins encoded by the sESts
CC   are useful in assays for determining biological activity and raising
CC   antibodies. They may be useful for treatment of autoimmune disorders
CC   (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC   (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC   osteoporosis, osteoarthritis, central nervous system disorders
CC   (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC   disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC   disease), tumours, bacterial, fungal or viral infections, depression and
CC   psoriasis. AAA45926 to AAA45931 represent linker variants which are given
CC   in the exemplification of the present invention.
XX
SQ   Sequence 1077 BP; 332 A; 160 C; 189 G; 396 T; 0 other;


alignment scores:
            Quality:    6.00              Length:      6
              Ratio:   1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_130_138 x AAA44456    ..

 Align seq 1/1  to: AAA44456  from: 1  to: 1077

        4 LysAspIleGluThrAla 8
          ||||||||||||||||||
      927 AAACACATAGAAACAGCA 944

seq name: /cgn2_6/prodata/2/genseq/seqdata/pCNA2000.DAT:AAC45629
```

```
Seq_documentation block:
ID   AAC35629 standard; DNA; 1078 BP
XX
AC   AAC35629;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 10867.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   07-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   24-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139764.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140824.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
```

PR 23 AUG 1999; 99US 0149902.
PR 23 AUG 1999; 99US 0149940.
PR 25 AUG 1999; 99US 0150566.
PR 26 AUG 1999; 99US 0150884.
PR 27 AUG 1999; 99US 0151065.
PR 27 AUG 1999; 99US-0151066.
PR 27 AUG 1999; 99US 0151080.
PR 30 AUG 1999; 99US 0151403.
PR 31 AUG 1999; 99US 0151448.
PR 01 SEP 1999; 99US 0151930.
PR 07 SEP 1999; 99US 0152363.
PR 10 SEP 1999; 99US-0153070.
PR 13 SEP 1999; 99US 0153758.
PR 15 SEP 1999; 99US 0154018.
PR 16 SEP 1999; 99US 0154039.
PR 20 SEP 1999; 99US-0154779.
PR 22 SEP 1999; 99US-0155139.
PR 23 SEP 1999; 99US-0155486.
PR 24 SEP 1999; 99US-0155659.
PR 28 SEP 1999; 99US-0156458.
PR 29 SEP 1999; 99US-0156596.
PR 04 OCT 1999; 99US 0157117.
PR 05 OCT 1999; 99US 0157754.
PR 06 OCT 1999; 99US-0157865.
PR 07 OCT 1999; 99US 0158029.
PR 08 OCT 1999; 99US 0158242.
PR 12 OCT 1999; 99US 0158869.
PR 13 OCT 1999; 99US 0159293.
PR 13 OCT 1999; 99US 0159294.
PR 13 OCT 1999; 99US 0159295.
PR 14 OCT 1999; 99US-0159329.
PR 14 OCT 1999; 99US 0159330.
PR 14 OCT 1999; 99US-0159331.
PR 14 OCT 1999; 99US 0159647.
PR 14 OCT 1999; 99US-0159648.
PR 18 OCT 1999; 99US 0159584.
PR 21 OCT 1999; 99US 0160741.
PR 21 OCT 1999; 99US-0160767.
PR 21 OCT 1999; 99US 0160768.
PR 21 OCT 1999; 99US 0160770.
PR 21 OCT 1999; 99US 0160814.
PR 21 OCT 1999; 99US-0160815.
PR 22 OCT 1999; 99US-0160980.
PR 22-OCT 1999; 99US 0160981.
PR 22 OCT-1999; 99US-0160989.
PR 25 OCT 1999; 99US-0161404.
PR 25 OCT 1999; 99US-0161405.
PR 25 OCT 1999; 99US 0161406.
PR 26 OCT-1999; 99US-0161359.
PR 26 OCT-1999; 99US-0161360.
PR 26 OCT-1999; 99US-0161361.
PR 28 OCT-1999; 99US-0161920.
PR 28 OCT 1999; 99US 0161992.
PR 28 OCT-1999; 99US 0161994.
PR 29 OCT 1999; 99US-0162142.

alignment scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2 (COPY_130_138) x AAC45629 ..

Align seq 1/1 to: AAC45629 from: 1 to: 1078

3 LysAspIleGluThrAla 8
||||||||||||||||||
550 AAAGACATAGAAACCGCT 567

```
OM nucleic - protein search, using frame_plus_p2n model
Run on: Sep 14 2001 09:53 PM

About: Results were produced by the GenCore software, version 4.5
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL frame+_p2n.model -DEV=xlp
 -Q [illegible]
 -DB Issued_Patents_NA -QFMT fastap -SUFFIX olip2n.rni
 -GAPOP=4.500 -GAPEXT=0.050 -MINMATCH 0.100 -LOOPCL=0.000
 -LOOPEXT 0.000 -QGAPOP=4.500 -QGAPEXT 0.050 -XGAPOP 60.000
 XGAPEXT 60.000 -FGAPOP=6.000 -FGAPEXT=7.000 YGAPOP 60.000
 -YGAPEXT 60.000 -DELOP 6.000 -DELEXT 7.000 START=1
 -MATRIX=oligo -TRANS=human40.cdi -LIST 45 -DOCALIGN=200
 -THR_SCORE quality -THR_MIN=1 ALIGN 15 MODE LOCAL OUTFMT pto
 -NORM ext -MINLEN=0 -MAXLEN 2000000000
 -USER US09126945_@CGN1_1_91 -NCPU 6 -ICPU 3 [illegible] NO_XLPY
 -WAIT -THREADS 1

Search information block:
query: US-09-126-945B-2_COPY_130_138
query length: 9
Database: Issued_Patents_NA:*
Database sequences: 324599
Database length: 94655562
Search time (sec): 322.720000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000


score_list:
Sequence                   Strd Orig   ZScore  E() [illegible]
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-841-646-3 +   6.00  102.69  419.83   17410  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-147-023-3 +   6.00  102.68  419.83   17410  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-447-570-3 +   6.00  102.69  419.83   17410  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-449-799-3 +   6.00  102.69  419.83   17410  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-449-[illegible] +   6.00  102.68  419.83   17410  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-[illegible] +   6.00  102.68  419.94   17125  !
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-07[illegible]   [illegible]  419.94  [illegible]
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-225-221-48   5.00  129.08  16.17   21  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-[illegible]-259-49   5.00  129.08  16.17   21  !
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:[illegible]-4169-19   5.00  129.08  16.17   21
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-[illegible]   [illegible]  [illegible]   [illegible]  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-[illegible]   [illegible]  [illegible]   [illegible]  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-[illegible]   5.00  109.42  199.69   [illegible]  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-[illegible]   5.00  [illegible]  [illegible]   [illegible]  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-[illegible]   5.00  109.05  211.09   [illegible]  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-[illegible]   5.00  107.76  219.86   [illegible]  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-714-918-6 +   5.00  107.52  225.71   412  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-265-315-6 +   5.00  107.52  225.71   412  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-265-315-6 +   5.00  107.52  225.71   412  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-266-417-6 +   5.00  107.52  225.71   412  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-470-179-73 +   5.00  107.22  231.56   422  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-470-179-150 +   5.00  107.31  232.05   [illegible]  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-068-756-617   5.00  107.21  232.05   [illegible]  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-680-506-10   5.00  106.94  243.21   446  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-388-267C-3 +   5.00  106.83  246.59   453  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-714-918-9 +   5.00  106.83  246.59   453  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-265-315-9 +   5.00  106.83  246.59   453  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-265-315-9 +   5.00  106.83  246.59   453  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-266-417-9 +   5.00  106.83  246.59   453  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-350-168-5 -   5.00  106.66  255.27   471  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-260-292A-5 +   5.00  104.37  338.31   647  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-017-114-9 +   5.00  104.37  338.31   647  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-505-207-9 +   5.00  104.37  338.31   647  !
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US94-02024-9 +   5.00  104.37  338.31   647
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US94-01149-25 +   5.00  104.01  351.04   681
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US94-01149-11 +   5.00  103.92  358.18   690
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US94-01149-71 +   5.00  103.92  358.19   690
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US94-01149-23 +   5.00  103.66  [illegible]   [illegible]
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-908-416-990 +   5.00  103.60  [illegible]   [illegible]
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US94-01149-17 +   5.00  103.54  376.09   [illegible]
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-908-416-[illegible] +   [illegible]  [illegible]  [illegible]   [illegible]
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-[illegible]-266-3 +   5.00  102.06  454.74   903  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-326-2PF-2 +   5.00  102.06  454.74   903  !
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US94-01149-31 +   5.00  101.99  458.76   9[illegible]
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US94-01149-56 +   5.00  101.97  460.10   9[illegible]

sequence: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-841-646-3

seq_documentation_block:
; Sequence 3, Application US/07841646
; Patent No. 5266683
;  GENERAL INFORMATION:
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  PANG, ROY H.L.
;    TITLE OF INVENTION:  OSTEOGENIC DEVICES
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  53 STATE STREET
;      CITY:  BOSTON
;      STATE:  MASSACHUSETTS
;      COUNTRY:  U.S.A.
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/841,646
;      FILING DATE:  19920221
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 810,560
;      FILING DATE:  20-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 827,052
;      FILING DATE:  28-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 660,162
;      FILING DATE:  22-FEB-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 621,988
;      FILING DATE:  04-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 621,849
;      FILING DATE:  04-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 616,374
;      FILING DATE:  21-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 600,024
;      FILING DATE:  18-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 599,543
;      FILING DATE:  18-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 579,865
;      FILING DATE:  07-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 569,920
;      FILING DATE:  20-AUG-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 483,913
;      FILING DATE:  22-FEB-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 422,613
;      FILING DATE:  17-OCT-1989
```

```
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 415,342
;       FILING DATE:  23-FEB-1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 232,630
;       FILING DATE:  15-AUG-1988
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 179,460
;       FILING DATE:  08-APR-1988
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  PITCHER, EDMUND R.
;       REGISTRATION NUMBER:  27,829
;       REFERENCE/DOCKET NUMBER:  CRP 001CP6
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617/248-7000
;       TELEFAX:  617/248-7100
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  17410 base pairs
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     ORIGINAL SOURCE:
;       ORGANISM:  homo sapiens
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  3192..3730
;       OTHER INFORMATION:  /label= EXON 1
;       OTHER INFORMATION:  /note= "START CODON BEGINS AT POSITION 3314"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10413..10414
;       OTHER INFORMATION:  /label= GAP 1
;       OTHER INFORMATION:  /note= "APPROXIMATELY _ BASES ARE ESTIMATED TO
;       OTHER INFORMATION:  BE MISSING BETWEEN POSITIONS 10413 AND 10414 IN
;       OTHER INFORMATION:  THIS SEQUENCE."
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  10696..10891
;       OTHER INFORMATION:  /label= EXON 2
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10960..10961
;       OTHER INFORMATION:  /label= GAP 2
;       OTHER INFORMATION:  /note= "APPROXIMATELY  BASES ARE ESTIMATED TO
;       OTHER INFORMATION:  BE MISSING BETWEEN POSITION 10960 AND 10961 IN
;       OTHER INFORMATION:  THIS SEQUENCE."
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  11059..11211
;       OTHER INFORMATION:  /label= EXON 3
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  11451..11452
;       OTHER INFORMATION:  /label= GAP 3
;       OTHER INFORMATION:  /note= "APPROXIMATELY  BASES ARE ESTIMATED TO
;       OTHER INFORMATION:  BE MISSING BETWEEN POSITIONS 11451 AND 11452 IN
;       OTHER INFORMATION:  THIS SEQUENCE."
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  11420..11617
;       OTHER INFORMATION:  /label= EXON 4
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  11721..11722
;       OTHER INFORMATION:  /label= GAP 4
;       OTHER INFORMATION:  /note= "APPROXIMATELY  BASES ARE ESTIMATED TO
;       OTHER INFORMATION:  BE MISSING BETWEEN POSITIONS 11721 AND 11722 IN
;       OTHER INFORMATION:  THIS SEQUENCE."
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  13354..13436
;       OTHER INFORMATION:  /label= EXON 5
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  15044..15160
;       OTHER INFORMATION:  /label= EXON 6
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  17245..17410
;       OTHER INFORMATION:  /label= EXON 7
US-07-841-646-3


Alignment scores:
                Quality:     6.00                Length:       6
                  Ratio:    1.000                  Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-126-945B-2-COPY_130_138 x US-07-841-646-3/rev  .

 Align Seq 1/.  to reverse of: US-07-841-646-3  from: 1  to: 17410

        2 LeuLysAspIleGluThr 7
          ||||||||||||||||||
    12276 CTAAAGACATAGAGACT 12260

 Seq Comm  exon  Seq Pat Inf_Seq_ID  exon  Seq US-08-147-023-4

Seq documentation block:
; Sequence 4, Application US/08147023
; Patent No. 5488045
; GENERAL INFORMATION:
;   APPLICANT:  OPPERMANN, HERMANN
;   APPLICANT:  OZKAYNAK, ENGIN
;   APPLICANT:  KUBEKACHEATH, THANGAVEL
;   APPLICANT:  RUEGER, DAVID C.
;   APPLICANT:  PANG, ROY H.L.
;   TITLE OF INVENTION:  OSTEOGENIC DEVICES
;   NUMBER OF SEQUENCES:  33
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;     STREET:  53 STATE STREET
;     CITY:  BOSTON
;     STATE:  MASSACHUSETTS
;     COUNTRY:  U.S.A.
;     ZIP:  02109
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC DOS/MS DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/147,023
;     FILING DATE:  21-FEB-1992
;     CLASSIFICATION:  530
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 810,560
;     FILING DATE:  20-DEC-1991
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 827,052
;     FILING DATE:  28-JAN-1992
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 660,162
;     FILING DATE:  22-FEB-1991
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 621,988
;     FILING DATE:  04-DEC-1990
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 621,849
;     FILING DATE:  04-DEC-1990
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 616,374
;     FILING DATE:  21-NOV-1990
```

```
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/447,924
;       FILING DATE:  18-OCT-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 599,543
;       FILING DATE:  18-OCT-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 579,865
;       FILING DATE:  07-SEP-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 569,920
;       FILING DATE:  20-AUG 1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 483,913
;       FILING DATE:  22-FEB-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 422,613
;       FILING DATE:  17-OCT-1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 315,342
;       FILING DATE:  23 FEB 1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 232,630
;       FILING DATE:  15-AUG-1988
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 179,460
;       FILING DATE:  08-APR-1988
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  PITCHER, EDMUND R.
;       REGISTRATION NUMBER:  27,829
;       REFERENCE/DOCKET NUMBER:  CRP-001CP6
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617/248 7000
;       TELEFAX:  617/248-7100
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  17410 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     ORIGINAL SOURCE:
;       ORGANISM:  homo sapiens
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  3192..3730
;       OTHER INFORMATION:  /label  EXON-1
;       OTHER INFORMATION:  /note  "START CODON BEGINS AT POSITION 3412"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10413..10414
;       OTHER INFORMATION:  /label  GAP-1
;       OTHER INFORMATION:  /note  "APPROXIMATELY _____ BASES ARE ESTIMATED T
;       OTHER INFORMATION:  BE MISSING BETWEEN POSITIONS 10413 AND 10414 IN
;       OTHER INFORMATION:  THIS SEQUENCE."
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  10696..10891
;       OTHER INFORMATION:  /label  EXON-2
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10960..10961
;       OTHER INFORMATION:  /label  GAP-2
;       OTHER INFORMATION:  /note  "APPROXIMATELY ____ BASES ARE ESTIMATED TO
;       OTHER INFORMATION:  BE MISSING BETWEEN POSITION 10960 AND 10961 IN
;       OTHER INFORMATION:  THIS SEQUENCE."
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  11059..11211
;       OTHER INFORMATION:  /label  EXON-3
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  11351..11352
;       OTHER INFORMATION:  /label  GAP-3
;       OTHER INFORMATION:  /note  "APPROXIMATELY ____ BASES ARE ESTIMATED TO
;       OTHER INFORMATION:  BE MISSING BETWEEN POSITIONS 11351 AND 11352 IN
;       OTHER INFORMATION:  THIS SEQUENCE."
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  11420..11617
;       OTHER INFORMATION:  /label  EXON-4
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  11721..11722
;       OTHER INFORMATION:  /label  GAP-4
;       OTHER INFORMATION:  /note  "APPROXIMATELY ____ BASES ARE ESTIMATED TO
;       OTHER INFORMATION:  BE MISSING BETWEEN POSITIONS 11721 AND 11722 IN
;       OTHER INFORMATION:  THIS SEQUENCE."
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  13854..14436
;       OTHER INFORMATION:  /label  EXON-5
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  15044..15160
;       OTHER INFORMATION:  /label  EXON-6
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  17245..17410
;       OTHER INFORMATION:  /label  EXON-7
US-08-147-024-3


alignment_scores:
                Quality:    6.00                Length:       6
                  Ratio:    1.000                 Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945b-2_COPY_130_138.x US-08-147-024-3/c      ..

 Align Seq 1/1  to reverse set: US-08-147-024-3  from: 1  to: 17410

        2 LeuLysAspIleGluThr 7
          ||||||||||||||||||
    12270 CTAAAAGACATAGAGACT 12253

Sequence 1 of 1 in Query Set 5A COMB Seq-US-08-447-570-3

seq_documentation_block:
; Sequence 3, Application US/08447570
; Patent No. 5714589
;  GENERAL INFORMATION:
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  CRAYNAE, ENCIN
;    AFFILIAN..  EUREKASAMPATH, THANCAVEL
;    APPLICANT:  SUESSER, DAVID C.
;    APPLICANT:  PANG, ROY H.L.
;    TITLE OF INVENTION:  OSTEOGENIC DEVICES
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  53 STATE STREET
;      CITY:  BOSTON
;      STATE:  MASSACHUSETTS
;      COUNTRY:  U.S.A.
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patentin Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/447,570
;      FILING DATE:  21-FEB-1992
;      CLASSIFICATION:  536
```

```
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 810,569
;       FILING DATE:  20-DEC-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 827,052
;       FILING DATE:  28-JAN-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 660,162
;       FILING DATE:  22-FEB-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 621,988
;       FILING DATE:  04-DEC-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 621,849
;       FILING DATE:  04-DEC-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 616,474
;       FILING DATE:  21-NOV-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 600,024
;       FILING DATE:  18-OCT-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 599,543
;       FILING DATE:  18-OCT-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 579,865
;       FILING DATE:  07-SEP-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 569,920
;       FILING DATE:  20-AUG-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 483,913
;       FILING DATE:  22-FEB-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 422,613
;       FILING DATE:  17-OCT-1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 315,442
;       FILING DATE:  23-FEB-1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 232,630
;       FILING DATE:  15-AUG-1988
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 179,460
;       FILING DATE:  08-APR-1988
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  PITCHER, EDMUND R.
;       REGISTRATION NUMBER:  27,829
;       REFERENCE/DOCKET NUMBER:  CRP 001CP6
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617/248-7000
;       TELEFAX:  617/248-7100
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  17410 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     ORIGINAL SOURCE:
;       ORGANISM:  homo sapiens
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  3192..3740
;       OTHER INFORMATION:  /label= EXON 1
;       OTHER INFORMATION:  /note= "START CODON BEGINS AT POSITION 3313"
;     FEATURE:
;       NAME/KEY:  misc feature
;       LOCATION:  10413..10414
;       OTHER INFORMATION:  /label= GAP 1
;       OTHER INFORMATION:  /note= "APPROXIMATELY       BASES ARE ESTIMATED TO
;       OTHER INFORMATION:  BE MISSING BETWEEN POSITIONS 10413 AND 10414 IN
;       OTHER INFORMATION:  THIS SEQUENCE."
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  10496..10891
;       OTHER INFORMATION:  /label= EXON 2
;     FEATURE:
;       NAME/KEY:  misc feature
;       LOCATION:  10960..10961
;       OTHER INFORMATION:  /label= GAP 2
;       OTHER INFORMATION:  /note= "APPROXIMATELY       BASES ARE ESTIMATED TO
;       OTHER INFORMATION:  BE MISSING BETWEEN POSITION 10960 AND 10961 IN
;       OTHER INFORMATION:  THIS SEQUENCE."
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  11059..11211
;       OTHER INFORMATION:  /label= EXON 3
;     FEATURE:
;       NAME/KEY:  misc feature
;       LOCATION:  11351..11352
;       OTHER INFORMATION:  /label= GAP 3
;       OTHER INFORMATION:  /note= "APPROXIMATELY       BASES ARE ESTIMATED TO
;       OTHER INFORMATION:  BE MISSING BETWEEN POSITIONS 11351 AND 11352 IN
;       OTHER INFORMATION:  THIS SEQUENCE."
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  11420..11617
;       OTHER INFORMATION:  /label= EXON 4
;     FEATURE:
;       NAME/KEY:  misc feature
;       LOCATION:  11721..11722
;       OTHER INFORMATION:  /label= GAP 4
;       OTHER INFORMATION:  /note= "APPROXIMATELY       BASES ARE ESTIMATED TO
;       OTHER INFORMATION:  BE MISSING BETWEEN POSITIONS 11721 AND 11722 IN
;       OTHER INFORMATION:  THIS SEQUENCE."
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  13854..14436
;       OTHER INFORMATION:  /label= EXON 5
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  15044..15160
;       OTHER INFORMATION:  /label= EXON 6
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  17245..17410
;       OTHER INFORMATION:  /label= EXON 7
; US-08-447-570-3


alignment_scores:
            Quality:    6,280              Length:       6
              Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_130_138 x US-08-447-570-3(rev    ..

  Align Seq 1/1   to reverse of: US-08-447-570-3   from: 1   to: 17410

        2 LeuLysAspIleGlnThr 7
          ||||||||||||||||
    12270 CTAAAAGACATAGAGACT 12253

seq1 name = us-09-126-945b-2_copy_130_138.seq.0, on 449,200 a

seq documentation_block:
; Sequence 3, Application US/08449700
; Patent No. 5865758
;  GENERAL INFORMATION:
;     APPLICANT:  GEFFERMANN, HERMANN
;     APPLICANT:  OZAYNAK, ENGIN
;     APPLICANT:  KUMARASAMPATH, THANGAVEL
;     APPLICANT:  RUEGER, DAVID C.
;     APPLICANT:  PANG, ROY H.L.
```

```
;     TITLE OF INVENTION:  OSTEOGENIC DEVICES
;     NUMBER OF SEQUENCES:  33
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  IESTA, BURWITZ & THIBEAULT
;       STREET:  53 STATE STREET
;       CITY:  BOSTON
;       STATE:  MASSACHUSETTS
;       COUNTRY:  U.S.A.
;       ZIP:  02109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/449,700
;       FILING DATE:  21-FEB-1992
;       CLASSIFICATION:  530
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 810,560
;       FILING DATE:  20-DEC-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 827,052
;       FILING DATE:  28-JAN-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 660,162
;       FILING DATE:  22-FEB-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 621,988
;       FILING DATE:  04-DEC-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 621,849
;       FILING DATE:  04-DEC-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 616,474
;       FILING DATE:  21-NOV-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 600,024
;       FILING DATE:  18-OCT-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 599,543
;       FILING DATE:  18-OCT-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 579,865
;       FILING DATE:  07-SEP-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 569,920
;       FILING DATE:  20-AUG-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 483,913
;       FILING DATE:  22-FEB-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 422,613
;       FILING DATE:  17-OCT-1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 315,342
;       FILING DATE:  24-FEB-1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 232,630
;       FILING DATE:  15-AUG-1988
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 179,460
;       FILING DATE:  08 APR 1988
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  PITCHER, EDMUND R.
;       REGISTRATION NUMBER:  27,829
;       REFERENCE/DOCKET NUMBER:  CRP 001CP6
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617/248-7000
;       TELEFAX:  617/248-7100
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  17410 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     ORIGINAL SOURCE:
;       ORGANISM:  homo sapiens
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  4192..4730
;       OTHER INFORMATION:  /label  EXON 1
;       OTHER INFORMATION:  /note  "START CODON BEGINS AT POSITION 441 "
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10413..10414
;       OTHER INFORMATION:  /label  GAP-1
;       OTHER INFORMATION:  /note  "APPROXIMATELY __ BASES ARE ESTIMATED TO
;       OTHER INFORMATION:  BE MISSING BETWEEN POSITIONS 10413 AND 10414 IN
;       OTHER INFORMATION:  THIS SEQUENCE."
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  10596..10891
;       OTHER INFORMATION:  /label  EXON 2
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10960..10961
;       OTHER INFORMATION:  /label  GAP-2
;       OTHER INFORMATION:  /note  "APPROXIMATELY ____ BASES ARE ESTIMATED TO
;       OTHER INFORMATION:  BE MISSING BETWEEN POSITION 10960 AND 10961 IN
;       OTHER INFORMATION:  THIS SEQUENCE."
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  11059..11211
;       OTHER INFORMATION:  /label  EXON-3
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  11351..11352
;       OTHER INFORMATION:  /label  GAP-3
;       OTHER INFORMATION:  /note  "APPROXIMATELY ____ BASES ARE ESTIMATED TO
;       OTHER INFORMATION:  BE MISSING BETWEEN POSITIONS 11351 AND 11352 IN
;       OTHER INFORMATION:  THIS SEQUENCE."
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  11420..11617
;       OTHER INFORMATION:  /label  EXON-4
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  11721..11722
;       OTHER INFORMATION:  /label  GAP-4
;       OTHER INFORMATION:  /note  "APPROXIMATELY ___ BASES ARE ESTIMATED TO
;       OTHER INFORMATION:  BE MISSING BETWEEN POSITIONS 11721 AND 11722 IN
;       OTHER INFORMATION:  THIS SEQUENCE."
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  13354..13446
;       OTHER INFORMATION:  /label  EXON-5
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  15044..15160
;       OTHER INFORMATION:  /label  EXON-6
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  17245..17410
;       OTHER INFORMATION:  /label  EXON-7
US 08/449,700-3


alignment_scores:
                Quality:    6.00              Length:       6
                  Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 Us 08/449,700-3  (US-08-449,700-3)  ..
```

```
 Align Seq 1/1  to reverse of: US-08-449-700A-3  from: 1  to: 17410

        2 LeuTyrAspIleGluThr 7
          TTTTTTTTTTTTTTTTTT
    12270 CTAAAAGACATAGAGACT 12253

Seq name: /cgn2_6/ptodata/2/fina/96-COMP.seq:US-09-449-699A-3

Seq documentation block:
; Sequence 3, Application US/08449699A
; Patent No. 5958441
;  GENERAL INFORMATION:
;     APPLICANT:  OPPERMANN, HERMANN
;     APPLICANT:  OZKAYNAK, ENGIN
;     APPLICANT:  KUBERASAMPATH, THANGAVEL
;     APPLICANT:  RUEGER, DAVID C.
;     APPLICANT:  PANG, ROY H.L.
;     TITLE OF INVENTION:  ANTIBODIES TO OSTEOGENIC PROTEINS
;     NUMBER OF SEQUENCES:  33
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;       STREET:  125 HIGH STREET
;       CITY:  Boston
;       STATE:  MASSACHUSETTS
;       COUNTRY:  U.S.A.
;       ZIP:  02110
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/449,699A
;       FILING DATE:  24-MAY-1995
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/147,023
;       FILING DATE:  01-NOV-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  PITCHER, EDMUND R.
;       REGISTRATION NUMBER:  27,829
;       REFERENCE/DOCKET NUMBER:  STR 001CP6CN
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617/248-7000
;       TELEFAX:  617/248-7100
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  17410 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     ORIGINAL SOURCE:
;       ORGANISM:  homo sapiens
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  3192..3740
;       OTHER INFORMATION:  /label= EXON 1
;       OTHER INFORMATION:  /note= "START OF 5'UN REGION AT -211 N 422"
;     FEATURE:
;       NAME/KEY:  misc feature
;       LOCATION:  10413..10414
;       OTHER INFORMATION:  /label= GAP 1
;       OTHER INFORMATION:  /note= "APPROXIMATELY       BASES ARE ESTIMATED I
;       OTHER INFORMATION:  BE MISSING BETWEEN POSITIONS 10413 AND 10414 IN
;       OTHER INFORMATION:  THIS SEQUENCE."
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  10696..10891
;       OTHER INFORMATION:  /label= EXON 2
;     FEATURE:
;       NAME/KEY:  misc feature
;       LOCATION:  10960..10961
;       OTHER INFORMATION:  /label= GAP 2
;       OTHER INFORMATION:  /note= "APPROXIMATELY       BASES ARE ESTIMATED I
;       OTHER INFORMATION:  BE MISSING BETWEEN POSITION 10960 AND 10961 IN
;       OTHER INFORMATION:  THIS SEQUENCE."
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  11059..11211
;       OTHER INFORMATION:  /label= EXON 3
;     FEATURE:
;       NAME/KEY:  misc feature
;       LOCATION:  11351..11352
;       OTHER INFORMATION:  /label= GAP 3
;       OTHER INFORMATION:  /note= "APPROXIMATELY       BASES ARE ESTIMATED I
;       OTHER INFORMATION:  BE MISSING BETWEEN POSITIONS 11351 AND 11352 IN
;       OTHER INFORMATION:  THIS SEQUENCE."
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  11420..11617
;       OTHER INFORMATION:  /label= EXON 4
;     FEATURE:
;       NAME/KEY:  misc feature
;       LOCATION:  11721..11722
;       OTHER INFORMATION:  /label= GAP 4
;       OTHER INFORMATION:  /note= "APPROXIMATELY       BASES ARE ESTIMATED I
;       OTHER INFORMATION:  BE MISSING BETWEEN POSITIONS 11721 AND 11722 IN
;       OTHER INFORMATION:  THIS SEQUENCE."
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  13354..13436
;       OTHER INFORMATION:  /label= EXON 5
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  15044..15160
;       OTHER INFORMATION:  /label= EXON 6
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  17245..17410
;       OTHER INFORMATION:  /label= EXON 7
US-08-449-699A-3


Alignment Statistics:
            Quality:    6.00              Length:       6
              Ratio:   1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-126-945B-2-COPY_130_138 x US-08-449-699A-3/rev  ..

 Align Seq 1/1  to reverse of: US-08-449-699A-3  from: 1  to: 17410

        2 LeuTyrAspIleGluThr 7
          TTTTTTTTTTTTTTTTTT
    12270 CTAAAAGACATAGAGACT 12253

Seq name: /cgn2_6/ptodata/2/fina/96-COMP.seq:US-09-449-699A-3

Seq documentation block:
; Sequence 3, Application US/08449699A
; Patent No. 5958441
;  GENERAL INFORMATION:
;     APPLICANT:  OZKAYNAK, ENGIN
;     APPLICANT:  OPPERMANN, HERMANN
;     TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR MODULATING
;     TITLE OF INVENTION:  MORPHOGENIC PROTEIN EXPRESSION
;     NUMBER OF SEQUENCES:  7
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
;       ADDRESSEE:  INC.
;       STREET:  45 SOUTH STREET
;       CITY:  HOPKINTON
;       STATE:  MA
```

```
;       COUNTRY:  USA
;       ZIP:  07148
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/486,343A
;       FILING DATE:  07-JUN-1995
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  PITCHER, Edmund R
;       REGISTRATION NUMBER:  27,829
;       REFERENCE/DOCKET NUMBER:  CRP-091CP
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617)-248-7000
;       TELEFAX:  (617)-248-7100
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  17415 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     ORIGINAL SOURCE:
;       ORGANISM:  homo sapiens
US-08-486-343A-1


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_130_138 x US-08-486-343A-1/c...

 Align seq 1/1  to reverse of: US-08-486-343A-1  from: 1  to: 17415

       2 LeuLysAspIleGluThr 7
         |||||||||||||||||
   12275 CTAAAACACATACACACT 12258

seq_name:  /tmp_1/pthaiba/2_fma/fCIUS_2.Mar26 [...] US95 07349 1

seq_documentation_block:
; Sequence 1, Application PC/TUS9507349
;  GENERAL INFORMATION:
;     APPLICANT:
;     TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR MODULATING
;     TITLE OF INVENTION:  MORPHOGEN EXPRESSION
;     NUMBER OF SEQUENCES:  7
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
;       ADDRESSEE:  INC
;       STREET:  45 SOUTH STREET
;       CITY:  HOPKINTON
;       STATE:  MA
;       COUNTRY:  USA
;       ZIP:  07148
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/07349
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/938,021
;       FILING DATE:  28-AUG-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  KELLEY, ROBIN D
;       REGISTRATION NUMBER:  34,637
;       REFERENCE/DOCKET NUMBER:  CRP-091PC
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (508) 435 9001
;       TELEFAX:  (508)-435-0992
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  17415 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (Genomic)
PCT-US95-07349-1


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_130_138 x PCT-US95-07349-1/seq...

 Align seq 1/1  to reverse of: PCT-US95-07349-1  from: 1  to: 17415

       2 LeuLysAspIleGluThr 7
         |||||||||||||||||
   12275 CTAAAACACATACACACT 12258

seq_name:  /tmp_1/pthaiba/2_fma/fCIUS_2.Mar26 [...] US95 07349 1

seq_documentation_block:
; Sequence 48, Application US/08225224
; Patent No. 5635599
;  GENERAL INFORMATION:
;     APPLICANT:  PASIAN, Ira
;     APPLICANT:  KREITMAN, Robert J.
;     TITLE OF INVENTION:  CIRCULARLY PERMUTATED LIGANDS ANI
;     TITLE OF INVENTION:  CIRCULARLY PERMUTED FUSION PROTEINS
;     NUMBER OF SEQUENCES:  57
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Townsend and Townsend Khourie and Crew
;       STREET:  Stewart Street Tower, One Market Plaza
;       CITY:  San Francisco
;       STATE:  California
;       COUNTRY:  US
;       ZIP:  94105-1494
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/225,224
;       FILING DATE:  8-APR 1994
;       CLASSIFICATION:  530
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Weber, Ellen L.
;       REGISTRATION NUMBER:  32,762
;       REFERENCE/DOCKET NUMBER:  15280 194
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 543-9600
;       TELEFAX:  (415) 543-5043
;   INFORMATION FOR SEQ ID NO:  48:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  21 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (primer)
US-08-225 224-48
```

```
alignment scores:
             Quality:     5.00              Length:       5
               Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_130_138 x US-08-225_224-48/rev    ..

 Align seq 1/1  to reverse of: US-08-225_224-48  from: 1  to: 21

        4 AspIleGluThrAla 8
          ||||||||||||||||
       18 GATAIAGAAACCGCG 4

Seq_name: /cgn2_6/ptodata/2/lma/6A_COMB.seq.US-08-722_258-48

Seq_documentation_block:
; Sequence 48, Application US/08722258
; Patent No. 6011002
;  GENERAL INFORMATION:
;     APPLICANT:  Pastan, Ira
;     APPLICANT:  Kreitman, Robert J.
;     APPLICANT:  Puri, Raj K.
;     TITLE OF INVENTION:  Circularly Permuted Ligands and
;     TITLE OF INVENTION:  Circularly Permuted Chimeric Molecules
;     NUMBER OF SEQUENCES:  72
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Townsend and Townsend and Crew LLP
;       STREET:  Two Embarcadero Center, Eighth Floor
;       CITY:  San Francisco
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  94111-3834
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/722,258
;       FILING DATE:  08-JAN-1997
;       CLASSIFICATION:  530
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  WO PCT/US95/04468
;       FILING DATE:  06-APR-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/225,224
;       FILING DATE:  08-APR-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Weber, Ellen Lauver
;       REGISTRATION NUMBER:  32,762
;       REFERENCE/DOCKET NUMBER:  015280-193-PCUS
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 576-0200
;       TELEFAX:  (415) 576-0300
;   INFORMATION FOR SEQ ID NO:  48:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  21 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA
;     FEATURE:
;       NAME/KEY:
;       LOCATION:  1..21
;       OTHER INFORMATION:  /note= "VK116 primer"
US-08-722-258-48


alignment scores:
             Quality:     5.00              Length:       5
               Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_130_138 x US-08-722-258-48/rev    ..

 Align seq 1/1  to reverse of: US-08-722-258-48  from: 1  to: 21

        4 AspIleGluThrAla 8
          ||||||||||||||||
       18 GATAIAGAAACCGCG 4

Seq_name: /cgn2_6/ptodata/2/lma/6A_COMB.seq.US-08-722-258-48

Seq_documentation_block:
; Sequence 48, Application PC/TUS9504468
;  GENERAL INFORMATION:
;     APPLICANT:
;     TITLE OF INVENTION:  CIRCULARLY PERMUTATED LIGANDS AND
;     TITLE OF INVENTION:  CIRCULARLY PERMUTED FUSION PROTEINS
;     NUMBER OF SEQUENCES:  59
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/04468
;       FILING DATE:  07-APR-1995
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/225,224
;       FILING DATE:  08-APR-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Weber, Ellen L.
;       REGISTRATION NUMBER:  32,762
;       REFERENCE/DOCKET NUMBER:  15280-194-IPC
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 543-9600
;       TELEFAX:  (415) 543-5043
;   INFORMATION FOR SEQ ID NO:  48:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  21 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (primer)
PCT-US95-04468-48


alignment scores:
             Quality:     5.00              Length:       5
               Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_130_138 x PCT-US95-04468-48/rev    ..

 Align seq 1/1  to reverse of: PCT-US95-04468-48  from: 1  to: 21

        4 AspIleGluThrAla 8
          ||||||||||||||||
       18 GATAIAGAAACCGCG 4

Seq_name: /cgn2_6/ptodata/2/lma/6A_COMB.seq.US-08-279_590A-3

Seq_documentation_block:
; Sequence 3, Application US/08279590A
; Patent No. 5656749
;  GENERAL INFORMATION:
;     APPLICANT:  ELIOT R. SPINDEL, SRINIVASA
;     APPLICANT:  NAGALLA AND BRENDA BARRY
;     TITLE OF INVENTION:  NUCLEIC ACIDS ENCODING
```

```
;     TITLE OF INVENTION:  RECEPTORS FOR BOMBESIN-LIKE
;     TITLE OF INVENTION:  PEPTIDES
;     NUMBER OF SEQUENCES:  10
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Fish & Richardson
;       STREET:  225 Franklin Street
;       CITY:  Boston
;       STATE:  Massachusetts
;       COUNTRY:  U.S.A.
;       ZIP:  02110-2804
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;       COMPUTER:  IBM PS/2 Model 50Z or 55SX
;       OPERATING SYSTEM:  MS-DOS (Version 5.0)
;       SOFTWARE:  WordPerfect (Version 5.1)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/279,590A
;       FILING DATE:  July 22, 1994
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  No. 5656749e
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Y. Rocky Tsao
;       REGISTRATION NUMBER:  34,053
;       REFERENCE/DOCKET NUMBER:  00537/200B001
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 542-5070
;       TELEFAX:  (617) 542-8906
;       TELEX:  200154
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  252
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-279-590A-3


alignment_scores:
            Quality:     5.00              Length:       5
              Ratio:    1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_130_138 x US-08-279-590A-3/rev   ..

 Align seq 1/1  to reverse of: US-08-279-590A-3  from: 1  to: 252

        1 ValLeuLysAspIle 5
          ||| |||||||||||
      140 GTTTTGAAGGACATA 126

seq_number  3 of /projdata2/lna/SA_COMP.seq for US-09-910-092-3

seq_documentation_block:
; Sequence 3, Application US/08610092
; Patent No. 5814463
; GENERAL INFORMATION:
;     APPLICANT:  Spindel, Eliot R.
;     APPLICANT:  Nagalla, Srinivasa
;     APPLICANT:  Barry, Brenda
;     TITLE OF INVENTION:  NUCLEIC ACIDS ENCODING RECEPTORS
;     TITLE OF INVENTION:  FOR BOMBESIN-LIKE PEPTIDES
;     NUMBER OF SEQUENCES:  10
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Fish & Richardson P.C.
;       STREET:  225 Franklin Street
;       CITY:  Boston
;       STATE:  MA
;       COUNTRY:  US
;       ZIP:  02110-2804
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  Windows95
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/910,092
;       FILING DATE:  12-AUG-1997
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/279,590
;       FILING DATE:  22-JUL-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Tsao, Y. Rocky
;       REGISTRATION NUMBER:  34,053
;       REFERENCE/DOCKET NUMBER:  00537/200B002
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617/542-5070
;       TELEFAX:  617/542-8906
;       TELEX:  200154
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  252 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     FEATURE:
;       NAME/KEY:  Coding Sequence
;       LOCATION:  1...252
US-08-910-092-3


alignment_scores:
            Quality:     5.00              Length:       5
              Ratio:    1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_130_138 x US-08-910-092-3/rev   ..

 Align seq 1/1  to reverse of: US-08-910-092-3  from: 1  to: 252

        1 ValLeuLysAspIle 5
          |||||||||||||||
      140 GTTTTGAAGGACATA 126

seq_number  1 of /projdata2/lna/SA_COMP.seq for US-09-060-756-451

seq_documentation_block:
; Sequence 451, Application US/09060756
; Patent No. 6183957
; GENERAL INFORMATION:
;  APPLICANT: Cole, Stewart
;  APPLICANT: Buchrieser-Brosch, Roland
;  APPLICANT: Gordon, Stephen
;  APPLICANT: billa..., Alain
;  TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
;  TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
;  TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
;  FILE REFERENCE: 8496-0169
;  CURRENT APPLICATION NUMBER: US/09/060,756
;  CURRENT FILING DATE: 1998-04-16
;  NUMBER OF SEQ ID NOS: 743
;  SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 451
;   LENGTH: 335
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (various positions within the sequence)
;   OTHER INFORMATION: applicants are uncertain of bases designated as "n"
US-09-060-756-451
```

```
alignment scores:
            Quality:     5.00              Length:       5
              Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2 (COPY_130_138) x US-09-060-256-251(rev   ..

 Align Seq 1/1  to reverse of: US-09-060-256-251  from: 1  to: 345

        5 IleGluThrAlaCys 9
          ||||||||||||||||
      278 ATCGAAACCGCCTGC 264

Seq name: /tmp/tmp/data/2/tmp/08_Mar_sq_1_07_pir.seq

Seq documentation block:
; Sequence 3, Application US/08259966
; Patent No. 5876985
;  GENERAL INFORMATION:
;     APPLICANT:  John F. Alderete
;     TITLE OF INVENTION:  Methods and Compositions for the Preparation of
;     TITLE OF INVENTION:  Recombinant Trichomonas Vaginalis Proteins and
;     TITLE OF INVENTION:  Peptides
;     NUMBER OF SEQUENCES:  4
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Arnold, White & Durkee
;       STREET:  P.O. Box 4433
;       CITY:  Houston
;       STATE:  Texas
;       COUNTRY:  USA
;       ZIP:  77210
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  WordPerfect 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/259,966
;       FILING DATE:  14 JUN 1994
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/692,982
;       FILING DATE:  25 April 1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Parker, David L.
;       REGISTRATION NUMBER:  32,165
;       REFERENCE/DOCKET NUMBER:  UTSK:134
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  512-320-7200
;       TELEFAX:  512 474 7577
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  339 base pairs [DNA of Figure 5]
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-259-966-3


alignment scores:
            Quality:     5.00              Length:       5
              Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2 (COPY_130_138) x US-08-259-966-3(rev   ..

 Align Seq 1/1  to reverse of: US-08-259-966-3  from: 1  to: 339

        5 IleGluThrAlaCys 9
          ||||||||||||||||
      29 ATTGAAACTGCCTGC 15
```

```
Seq name: /tmp/tmp/data/2/tmp/08_Mar_sq_1_07_pir.seq.US-08-701-234B-2

Seq documentation block:
; Sequence 2, Application US/08701234B
; Patent No. 5861408
;  GENERAL INFORMATION:
;     APPLICANT:  Pfroundschuh, Michael; Renner, Christoph
;     TITLE OF INVENTION:  ISOLATED NUCLEIC ACIDS ASSOCIATED WITH T CELL
;     TITLE OF INVENTION:  ACTIVATION AND USES THEREOF
;     NUMBER OF SEQUENCES:  2
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Foley & Lynch
;       STREET:  805 Third Avenue
;       CITY:  New York City
;       STATE:  New York
;       ZIP:  10022
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 3.5 inch, 360 kb storage
;       COMPUTER:  IBM
;       OPERATING SYSTEM:  PC-DOS
;       SOFTWARE:  Wordperfect
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/701,234B
;       FILING DATE:  21 August 1996
;       CLASSIFICATION:  536
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Hanson, No. 5861408man D.
;       REGISTRATION NUMBER:  30,946
;       REFERENCE/DOCKET NUMBER:  LUD 5440
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 688 9200
;       TELEFAX:  (212) 838-3884
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  380 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-701-234B-2


alignment scores:
            Quality:     5.00              Length:       5
              Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2 (COPY_130_138) x US-08-701-234B-2(rev   ..

 Align Seq 1/1  to reverse of: US-08-701-234B-2  from: 1  to: 380

        1 ValGlnLysAspIle 5
          |||||||||||||||
      54 GTTTGAAAGACATT 40
```

```
OM of: US-09-126-945B-2_COPY_130_138 to: EST:*   out_format : pts

Date: Sep 14, 2001  7:47 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL frame+_p2n.model -DEV=xlp
 -Q [illegible] 1267
 -DB=EST -QFMT=fastap -SUFFIX=olip2n.rst -GAPOP=4.500
 -GAPEXT 0.050 -MINMATCH 0.100 -LOOPCL 0.000 -LOOPEXT 0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT 60.000
 -FGAPOP 6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT 60.000
 -DELOP=6.000 -DELEXT 7.000 -START=1 -MATRIX=oligo
 -TRANS human40.cdi -LIST=45 -DOCALIGN 200 -THR_SCORE quality
 -THR_MIN 1 -ALIGN 15 -MODE=LOCAL -OUTFMT pfs -NORM ext -MINLEN 0
 -MAXLEN=2000000000 -USER=US09126945_@CGN1_1_1515 -NCPU 6
 -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-126-945B-2_COPY_130_138
Query length: 9
Database: EST:*
Database sequences: 10228115
Database length: 431459454
Search time (sec): 9087.630000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000


score_list:
Sequence               Strd Orig     ZScore   EScore Len  ! Documentation  ..
gb_est78:BE757615       +     9.00   200.44   0.0479   539  ! BE757615 212104 MARC 2BOV Bos t
gb_est97:BG179568       +     9.00   197.14   0.0732   991  ! BG179568 602328182F1 NIH_MGC_9
gb_est89:BF579116       +     9.00   193.16   0.1219   2064 ! BF579116 602096050F1 NCI_CGAP_
gb_gss14:AQ628603       -     7.00   156.70   13.09    425  ! AQ628603 CITBI E1 2654E14.TR C
gb_gss11:AQ443287       +     7.00   156.50   13.43    441  ! AQ443287 HS_5050_A2_H05_SP6E RP
gb_est22:AI594599       +     7.00   155.87   14.56    495  ! AI594599 vo79e01.y1 Barstead mo
gb_est9:AA615554        +     7.00   155.55   15.17    525  ! AA615554 vo79e01.r1 Barstead mo
gb_gss12:AQ490676       +     7.00   154.76   16.79    608  ! AQ490676 RPCI-11-268B3.TJ RPCI
gb_est89:BF535871       +     7.00   154.50   17.35    637  ! BF535871 602051384F1 NCI_CGAP_S
gb_est99:BG327946       +     7.00   151.73   24.77    1063 ! BG327946 602426969F1 NIH_MGC_1
gb_gss29:AZ586687             6.00   149.76   31.97    24   ! AZ586687 1M0352F10F Mouse 10kb p
gb_est27:AI959857       +     6.00   141.04   97.58    120  ! AI959857 sc34e06.x1 Gm c1014 CE
gb_est4:AA279099        -     6.00   140.48   104.81   133  ! AA279099 zs82e06.s1 NCI_CGAP_GC
gb_gss12:AQ479538       +     6.00   138.78   130.35   182  ! AQ479538 RPCI-11-232E19.TJ RPCI
gb_gss37:TA152A10P      +     6.00   138.21   140.15   202  ! AL467175 T. brucei sheared gen
gb_est59:BB201357       +     6.00   138.13   141.59   205  ! BB201357 BB201357 RIKEN full-le
gb_est90:BF612753       -     6.00   138.13   141.59   205  ! BF612753 dd76a11.x2 Wellcome CR
gb_est106:N75663        +     6.00   137.32   157.08   238  ! N75663 yv29d04.r1 Soares fetal
gb_gss35:B75106         -     6.00   137.32   157.08   238  ! B75106 RPCI11 13F6.TV RPCI 11 H
gb_est8:AA506674        +     6.00   137.08   162.09   249  ! AA506674 nb47f08.s1 NCI_CGAP_Pr
gb_est33:AV130123       +     6.00   136.93   165.24   256  ! AV130123 AV130123 Mus musculus
gb_est25:AI827091       +     6.00   136.38   177.17   283  ! AI827091 wk89h01.x1 NCI_CGAP_La
gb_est98:BG237482       +     6.00   136.07   184.51   300  ! BG237482 sab07b09.y1 Gm c1071 s
gb_est68:BB565019       +     6.00   136.01   185.79   303  ! BB565019 BB565019 RIKEN full-le
gb_est5:AA303498        +     6.00   135.96   187.06   306  ! AA303498 EST16126 Aorta endothe
gb_est63:BB361943       +     6.00   135.96   187.06   306  ! BB361943 BB361943 RIKEN full le
gb_est2:AA091488        -     6.00   135.93   187.91   308  ! AA091488 113510.seq.F Human fet
gb_est57:BB129495       +     6.00   135.91   188.34   309  ! BB129495 BB129495 RIKEN full-le
gb_est77:BE625773       -     6.00   135.79   191.29   316  ! BE625773 ut98h07.y2 Soares_mamm
gb_est9:AA566674        +     6.00   135.77   191.71   317  ! AA566674 ZF-A258 Zebrafish adul
gb_est107:R01826        -     6.00   135.65   194.65   324  ! R01826 ye84f06.r1 Soares fetal
gb_gss19:AQ983727       -     6.00   135.57   196.73   329  ! AQ983727 RPCI-23-322A2.TV RPCI-
gb_est54:AW988548       -     6.00   135.39   201.28   340  ! AW988548 uq08g11.y1 Soares_mamm
gb_est12:AA793291       -     6.00   135.28   204.15   347  ! AA793291 vm12h08.r1 Knowles Sol
gb_est12:AA785377             6.00   135.25   204.97   349  ! AA785377 q6h04a1.t1 Aspergillus
gb_est75:BE522462       -     6.00   135.11   208.63   358  ! BE522462 M26A8CIM Arabidopsis c
gb_est34:AV194158       +     6.00   135.08   209.44   360  ! AV194158 AV194158 Yuji Kohara u
gb_est103:C08723        +     6.00   135.08   209.44   360  ! C08723 C08723 Yuji Kohara unpub
gb_est103:C64245        +     6.00   135.08   209.44   360  ! C64245 C64245 Yuji Kohara unpub
gb_est23:AI661874       +     6.00   135.01   211.46   365  ! AI661874 mt55e10.x1 Barstead
gb_gss11:AQ451600       -     6.00   134.95   213.07   369  ! AQ451600 HS_5157_A1_E02_SP6E
gb_gss7:AQ099967        +     6.00   134.83   216.27   377  ! AQ099967 HS_3054_B1_B10_MF C
gb_est27:AJ237164       +     6.00   134.74   218.66   383  ! AJ237164 AJ237164 Emericella
gb_est109:T54021        -     6.00   134.65   221.43   390  ! T54021 yb85f09.r1 Stratagene
gb_est22:AI595643             6.00   134.60   222.61   393  ! AI595643 ac11b01.x1 Soares m

seq_name: gb_est78:BE757615

seq_documentation_block:
LOCUS       BE757615        539 bp    mRNA            EST       15-SEP-2000
DEFINITION  212104 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BE757615
VERSION     BE757615.1  GI:10171607
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 539)
  AUTHORS   Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
            Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
            ,W.W. and Keele,J.W.
  TITLE     Design and use of four pooled tissue normalized cDNA libraries for
            EST discovery in cattle
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the  minscore 18
            and  minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 64  row: o  column: 13
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1..539
                     /organism "Bos taurus"
                     /db_xref "taxon:9913"
                     /clone_lib "MARC 2BOV"
                     /tissue_type "pooled"
                     /lab_host "DH10B"
                     /note "Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from testis, thymus,
                     semitendonosus muscle, longissimus muscle, pancreas,
                     adrenal, and endometrium."
BASE COUNT        98 a    168 c    193 g     80 t
ORIGIN


alignment_scores:
          Quality:     9.00              Length:        9
            Ratio:   1.000                 Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_130_138 x BE757615   ..

 Align seq 1/1  to: BE757615  from: 1  to: 539

      1 ValLeuLysAspIleGluThrAlaCys 9
        |||||||||||||||||||||||||||
    140 GTGCTCAAGGACATCGAGACGGCCTGC 166

seq_name: gb_est97:BG179568

seq_documentation_block:
```

```
LOCUS       BG179568       991 bp    mRNA            EST       06 FEB 2001
DEFINITION  602428182F1 NIH_MGC 91 Homo sapiens cDNA clone IMAGE:4429413 5',
            mRNA sequence.
ACCESSION   BG179568
VERSION     BG179568.1  GI:12686271
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 991)
  AUTHORS   NIH_MGC http://mgc.nci.nih.gov/
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/TMP
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10181  row: d  column: 22
            High quality sequence stop: 655.
FEATURES             Location/Qualifiers
     source          1..991
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4429413"
                     /clone_lib="NIH_MGC_91"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: prostate; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo dT primed.
                     Average insert size 1.4 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      190 a    351 c    305 g    145 t
ORIGIN


Alignment scores:
            Quality:     9.00            Length:       9
              Ratio:    1.000              Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

Alignment block:
 US-09-126-945B-2 COPY_130_138 x BG179568    ..

 Align seq 1/1  to: BG179568  from: 1  to: 991

        1 ValLeuLysAspIleGluThrAlaCys 9
          ||||||||||||||||||||||||||||
      291 GTGCTCAAGGACATCGAGACGGCCTGC 317

Seq_name: gb_est89:BF579116

Seq_documentation_block:
LOCUS       BF579116      2064 bp    mRNA            EST       12-DEC-2000
DEFINITION  602096050F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4215901 5',
            mRNA sequence.
ACCESSION   BF579116
VERSION     BF579116.1  GI:11652828
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2064)
  AUTHORS   NIH_MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM09292  row: d  column: 14
            High quality sequence stop: 645.
FEATURES             Location/Qualifiers
     source          1..2064
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4215901"
                     /clone_lib="NCI_CGAP_Co24"
                     /lab_host="DH10B (T1 phage resistant)"
                     /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: oligo dT.
                     Average insert size 1.6 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      624 a    621 c    620 g    199 t
ORIGIN


Alignment scores:
            Quality:     9.00            Length:       9
              Ratio:    1.000              Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

Alignment block:
 US-09-126-945B-2 COPY_130_138 x BF579116    ..

 Align seq 1/1  to: BF579116  from: 1  to: 2064

        1 ValLeuLysAspIleGluThrAlaCys 9
          ||||||||||||||||||||||||||||
      344 GTGCTCAAGGACATCGAGACGGCCTGC 370

Seq_name: gb_gss14:AQ628603

Seq_documentation_block:
LOCUS       AQ628603       425 bp    DNA             GSS       16-JUN-1999
DEFINITION  CITBI-E1-2654E14.1R CITBI-E1 Homo sapiens genomic clone 2654E14,
            DNA sequence.
ACCESSION   AQ628603
VERSION     AQ628603.1  GI:5090995
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 425)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from Caltech Libraries for Sequence Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbmaint.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/mdb/bacfos/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
```

```
     source          1. .425
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone "2654E14"
                     /clone_lib="CITBI-E1"
                     /sex="male"
                     /cell_type="sperm"
                     /note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
                     CalTech Human BAC Library D"
BASE COUNT       134 a      72 c      99 g     120 t
ORIGIN


alignment_scores:
            Quality:     7.00              Length:       7
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_130_138 x AQ628603/rev    ..

 Align seq 1/1  to reverse of: AQ628603  from: 1  to: 425

       2 LeuLysAspIleGluThrAla 8
         ||||||||||||||||||||||
     266 TTGAAGGACATTGAGACTGCA 246

seq_name: qb_qss11:AQ443287

seq_documentation_block:
LOCUS        AQ443287      441 bp     DNA              GSS        31 MAR 1999
DEFINITION   HS_5050_A2_H05_SP6E RPCI-11 Human Male BAC Library Homo sapiens
             genomic clone Plate=626 Col=10 Row=O, DNA sequence.
ACCESSION    AQ443287
VERSION      AQ443287.1  GI:4554626
KEYWORDS     GSS.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 441)
  AUTHORS    Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
             Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
             Hood,L.
  TITLE      Sequence-tagged connectors: A sequence approach to mapping and
             scanning the human genome
  JOURNAL    Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE    99380589
COMMENT      Contact: Mahairas GG, Wallace JC, Hood L
             High Throughput Sequencing Center
             University of Washington
             401 Queen Anne Avenue North, Seattle, WA 98109, USA
             Tel: (206) 616-3618
             Fax: (206) 616-3887
             Email: jwallace@u.washington.edu
             Clones are derived from the human BAC library RPCI-11. For BAC
             library availability, please contact Pieter de Jong
             (pieter@dejong.med.buffalo.edu). Clones may be purchased from
             BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
             or from Resear h Genetics (info@resgen.com). BAC end Web Server:
             http://www.hgsc.washington.edu
             Plate: 626  row: O  column: 10
             Seq primer: SP6
             Class: BAC ends
             High quality sequence stop: 411
FEATURES             Location/Qualifiers
     source          1. .441
                     /organism= "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone="Plate=626 Col=10 Row=O"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT       121 a     103 c      93 g     124 t
ORIGIN


alignment_scores:
            Quality:     7.00              Length:       7
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_130_138 x AQ443287    ..

 Align seq 1/1  to: AQ443287  from: 1  to: 441

       2 LeuLysAspIleGluThrAla 8
         ||||||||||||||||||||||
     320 CTGAAAGATATAGAAACTGCA 340

seq_name: qb_est22:A1594599

seq_documentation_block:
LOCUS        A1594599      495 bp     mRNA             EST        21-APR-1999
DEFINITION   vo79c01.y1 Barstead mouse myotubes MPLRB5 Mus musculus cDNA clone
             IMAGE:1065312 5', mRNA sequence.
ACCESSION    A1594599
VERSION      A1594599.1  GI:4603647
KEYWORDS     EST.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 495)
  AUTHORS    Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
             Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
             ,P., Sweller,T., Gibbens,M., Pape,D., Harvey,N., Schurk,E., Ritter
             ,E., Kohn,S., Shin,I., Jackson,Y., Cardenas,M., McCann,K.,
             Waterston,R. and Wilson,R.
  TITLE      The WashU-NCI Mouse EST Project 1999
  JOURNAL    Unpublished (1999)
COMMENT      Contact: Marra M/WashU-NCI Mouse EST Project 1999
             Washington University School of Medicine
             4444 Forest Park Parkway, Box 8501, St. Louis, Mo 64108, USA
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: mouseest@watson.wustl.edu
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             MGI:587672
             This read is a RESEQUENCE of a previously sequenced mouse clone
             This read has been verified (found to hit its original self in the
             correct orientation)
             Seq primer:  40RP from Gibco
             High quality sequence stop: 440.
FEATURES             Location/Qualifiers
     source          1   495
                     /organism="Mus musculus"
                     /strain="C3H"
                     /db_xref "taxon:10090"
                     /clone "IMAGE:1065312"
                     /clone_lib "Barstead mouse myotubes MPLRB5"
                     /cell_line "C2C12"
                     /lab_host "DH10B"
                     /note "Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: Not1; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACGAATCTGAAGTGGGAGCGGCCGCCTTTTTTTTTTTTTTTTTT3'];  double stranded cDNA was ligated to Eco RI adaptors
                     [AATTCGGATCCTTG], digested with Not I and cloned into the
                     Not I  and Eco RI sites of the modified pT7T3 vector.
```

```
                    Library constructed by Bob Barstead.  The C2C12 cell line
                    (available from ATCC, catalog # CRL 1772) differentiates
                    rapidly, forming contractile myotubes and producing
                    characteristic muscle proteins.  "
BASE COUNT       145 a    103 c     98 g    149 t
ORIGIN


alignment scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_130_138 x AI594599    ..

 Align seq 1/1  to: AI594599  from: 1  to: 495

        1 ValLeuLysAspIleGluThr 7
          |||||||||||||||||||
      167 GTCTTAAAAGAC'ATAGAAAC' 187

Seq name: gb_est9:AA615554

Seq documentation block:
LOCUS        AA615554       525 bp     mRNA             EST        07-OCT-1997
DEFINITION   vo79e01.r1 Barstead mouse myotubes MPLRB5 Mus musculus cDNA clone
             IMAGE:1065412 5', mRNA sequence.
ACCESSION    AA615554
VERSION      AA615554.1  GI:2502782
KEYWORDS     EST.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 525)
  AUTHORS    Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
             Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
             Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
             Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
             Waterston,R.
  TITLE      The Wash0-HHMI Mouse EST Project
  JOURNAL    Unpublished (1996)
COMMENT      Contact: Marra M/Mouse EST Project
             Wash0 HHMI Mouse EST Project
             Washington University School of Medicine
             4444 Forest Park Parkway, Box 8501, St. Louis, Mo 63108
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: mouseest@watson.wustl.edu
             This clone is available royalty free through LLNL; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             MGI:587672
             Seq primer:  28ml3 rev2 ET from Amersham
             High quality sequence stop: 517.
FEATURES             Location/qualifiers
     source          1..525
                     /organism="Mus musculus"
                     /strain="C3H"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1065412"
                     /clone_lib="Barstead mouse myotubes MPLRB5"
                     /cell_line="C2C12"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
                     was primed with a Not I  oligo(dT) primer (5'
                     TGTTACGAATTCTGAAGTGGGAGGGGCGGCCGCTTTTTTTTTTTTTTTTTTTTTTTT
                     3'); double-stranded cDNA was ligated to Eco RI adaptors
                     [AATTCGGATCCTTG], digested with Not I and cloned into the
                     Not I  and Eco RI sites of the modified pT7T3 vector.
                     Library constructed  by Bob Barstead.  The C2C12 cell line
                     (available from ATCC, catalog # CRL 1772) differentiates
                     rapidly, forming contractile myotubes and producing
                     characteristic muscle proteins.  "
BASE COUNT       152 a    110 c    105 g    158 t
ORIGIN


alignment scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_130_138 x AA615554    ..

 Align seq 1/1  to: AA615554  from: 1  to: 525

        1 ValLeuLysAspIleGluThr 7
          |||||||||||||||||||
      165 GTCTTAAAAGAC'ATAGAAAC' 185

Seq name: gb_gss12:AQ490676

Seq documentation block:
LOCUS        AQ490676       608 bp     DNA              GSS        24-APR-1999
DEFINITION   RPCI-11-268B3.TV RPCI-11 Homo sapiens genomic clone RPCI-11-268B3,
             DNA sequence.
ACCESSION    AQ490676
VERSION      AQ490676.1  GI:4676550
KEYWORDS     GSS.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 608)
  AUTHORS    Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
             ,J.C.
  TITLE      Use of BAC End Sequences from Library RPCI-11 for Sequence Ready
             Map Building
  JOURNAL    Unpublished (1997)
COMMENT      Other_GSSs: RPCI-11-268B3.TV
             Contact: Shaying Zhao, William Nierman, Mark Adams
             Department of Eukaryotic Genomics
             The Institute for Genomic Research
             9712 Medical Center Dr., Rockville, MD 20850
             Tel: 301 838 0200
             Fax: 301 838 0208
             Email: hbac@tigr.org
             Clones are derived from the human BAC library RPCI-11. For BAC
             library availability, please contact Pieter de Jong
             (pieter@dejong.med.buffalo.edu). Clones may be purchased from
             BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
             Research Genetics (info@resgen.com). BAC end search page:
             http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
             Seq primer: SP6
             Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..608
                     /organism="Homo sapiens"
                     /db_xref="GDB:7602554"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11 268B3"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI-11 Human Male BAC Library"
BASE COUNT       187 a    144 c    117 g    158 t      2 others
ORIGIN


alignment scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000               Gaps:       0
```

```
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_130_138 x AQ490676     .

 Align seq 1/1  to: AQ490676  from: 1  to: 608

        2 LeuLysAspIleGluThrAla 8
          ||||||||||||||||||||
      111 CTGAAAGACATAGAAACTGCT 131

seq_name: qb_est89:BF535871

seq_documentation_block:
LOCUS        BF535871       637 bp    mRNA            EST       11-DEC-2000
DEFINITION  602051384F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4190637 5',
            mRNA sequence.
ACCESSION   BF535871
VERSION     BF535871.1  GI:11624289
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 637)
  AUTHORS   NIH MGC Project
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9518  row: q  column: 22
            High quality sequence stop: 637.
FEATURES             Location/Qualifiers
     source          1..637
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4190637"
                     /clone_lib="NCI_CGAP_SG2"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: salivary gland; Vector: pCMV-SPORT6; Site_1:
                     NotI; Site_2: SalI; Cloned unidirectionally. Primer: oligo
                     dT. Average insert size 1.8 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT       120 a    219 c    179 g    114 t      5 others
ORIGIN


alignment_scores:
            Quality:    7.00            Length:      7
              Ratio:   1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_130_138 x BF535871    ..

 Align seq 1/1  to: BF535871  from: 1  to: 637

        1 ValLeuLysAspIleGlnThr 7
          |||||||||||||||||||
      616 GTCCTGAAAGATATTGAGACG 636

seq_name: qb_est99:BG327946

seq_documentation_block:
LOCUS        BG327946      1063 bp    mRNA            EST       27-FEB-2001
DEFINITION  602426969F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564720 5',
            mRNA sequence.
ACCESSION   BG327946
VERSION     BG327946.1  GI:13134384
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1063)
  AUTHORS   NIH MGC Project
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue procurement: UCSD/ITP
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LICM1280  row: j  column: 17
            High quality sequence stop: 713.
FEATURES             Location/Qualifiers
     source          1..1063
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4564720"
                     /clone_lib="NIH_MGC_14"
                     /tissue_type="renal cell adenocarcinoma"
                     /lab_host="DH10B (phage resistant)"
                     /note="Organ: kidney; Vector: pcDNA2.1; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP cDNA synthesis kit
                     (Stratagene) and Superscript II RI (Life Technologies)."
BASE COUNT       344 a    253 c    258 g    208 t
ORIGIN


alignment_scores:
            Quality:    7.00            Length:      7
              Ratio:   1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_130_138 x BG327946    ..

 Align seq 1/1  to: BG327946  from: 1  to: 1063

        3 LysAspIleGluThrAlaCys 9
          |||||||||||||||||||
      596 AAAGACATAGAAACGGCTTGC 616

seq_name: qb_gss29:AZ586687

seq_documentation_block:
LOCUS        AZ586687       24 bp    DNA             GSS       13-DEC-2000
DEFINITION  IM0392K16R Mouse 10kb plasmid UUGClM Library Mus musculus genomic
            clone UUGClM0392K16 R, DNA sequence.
ACCESSION   AZ586687
VERSION     AZ586687.1  GI:11708877
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 24)
  AUTHORS   Dunford,J., Ayyagi,A., Barbet,M., Beaconnais,J., Bevan,S., Bonnet,C.,
```

```
            Islam,H., Longacre,S., Mahmoud,M., Meetenu,E., Pederson,T., Reilly
            ,M., Rose,M., Rose,E., Stokes,K., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000  Std Error: 0.00
            Plate: 0492  row: K  column: 16
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 24.
FEATURES             Location/Qualifiers
     source          1..24
                     /organism "Mus musculus"
                     /strain "C57BL/6J"
                     /db_xref "taxon:10090"
                     /clone "UUEC1M0492K16"
                     /clone_lib "Mouse 10kb plasmid UUGC1M library"
                     /sex "Male"
                     /lab_host "E. Coli strain XL10-Gold, Tl resistant, F'"
                     /note "Vector: PWD4,cos: Partially prepared DNA from M
                     musculus C57BL/6J (male) was obtained from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/cmdr). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4742114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically competent E. coli XL10 Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT        5 a        3 c        7 g        9 t
ORIGIN


alignment scores:
            Quality:     6.00              Length:       6
              Ratio:    1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY-130-138 x AZ586687/rc  .

 Align seq 1/1  1 ccccccca  8  AZ586687  1 c    1  21

        2 IleLysAspIleGluThr 7
          |||||||||||||||||
       21 CTCAAAGACATTGAGACT 4

Seq name: gb:est27:AI959857

Seq documentation block:
LOCUS       AI959857       120 bp    mRNA            EST       20-AUG-1999
DEFINITION  zc84c06.y1 Gm-cl014 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-cl014-995 3', mRNA sequence.
ACCESSION   AI959857
VERSION     AI959857.1  GI:5752570
```

```
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE   1  (bases 1 to 120)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Cregell,V., Khanna
            ,A., Bella,R., Marra,M., Billier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,L., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, Mo 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (888) 430-0030 or (314) 427-5222 FAX:(888) 919-5321 or (314)
            427-5424 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            Seq primer: primer name ambiguous.
FEATURES             Location/Qualifiers
     source          1..120
                     /organism "Glycine max"
                     /db_xref "taxon:3847"
                     /clone "GENOME SYSTEMS CLONE ID: Gm-cl014-995"
                     /clone_lib "Gm-cl014"
                     /tissue_type "Leaves, 2-3 week old seedlings, greenhouse
                     grown"
                     /lab_host "DH10B"
                     /note "Vector: pT7T3Pac (pT7T3, Pharmacia); Site_1: EcoRI;
                     Site_2: HindIII; This cDNA library was constructed from
                     mRNA isolated from leaves of 2-3 week old greenhouse grown
                     plants. Complementary DNA was synthesized from mRNA using
                     a 3' anchored poly (dT) primer. EcoRI adaptors were
                     ligated to the blunt-ended cDNA fragments followed by
                     digestion with EcoRI and HindIII. The cDNA fragments were
                     directionally cloned into the EcoRI-HindIII restriction
                     site of the pT7T3-Pac vector. The ligated cDNA fragments
                     were transformed into DH10B host cells (Gibco BRL). This
                     library was constructed by Dr. Randy Shoemaker and Dr.
                     John Erpelding."
BASE COUNT       40 a       19 c       27 g       34 t
ORIGIN


alignment scores:
            Quality:     6.00              Length:       6
              Ratio:    1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY-130-138 x AI959857  ..

 Align seq 1/1  to: AI959857  from: 1  to: 120

       1 ValLeuLysAspIleGlu 6
         |||||||||||||||||
      12 GTCTCAAAGACATTGAG 29

Seq name: gb:est14:AA279099

Seq documentation block:
LOCUS       AA279099       133 bp    mRNA            EST       15-AUG-1997
DEFINITION  zs82e06.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:294002 3'
            similar to contains L1.t3 L1 repetitive element ;, mRNA sequence.
```

```
ACCESSION   AA279099
VERSION     AA279099.1  GI:1920564
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 133)
  AUTHORS   NCI CGAP http://www.ncbi.nlm.nih.gov/ncicgap
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through the I.M.A.G.E.
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: 41m13 fwd  PT ::= AccI,HincII
            High quality sequence stop: 111.
FEATURES             Location/Qualifiers
     source          1..133
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:704002"
                     /clone_lib="NCI_CGAP_GC1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3-Pac (Pharmacia); with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCICA11111111111111111-3'
                     ]. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT        28 a      23 c      16 g      66 t
ORIGIN


alignment_scores:
            Quality:    6.00            Length:      6
              Ratio:   1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_130_138 x AA279099/rev    .

 Align seq 1/1  to reverse of AA279099  from: 1  to: 133

       2 LeuLysAspIleGluThr 7
         |||||||||||||||||||
     101 CTGAAGGATATTGAAACA 84

Seq_name: qb_qss12:AQ479538

seq_documentation_block:
LOCUS       AQ479538      182 bp    DNA             GSS       23 APR 1999
DEFINITION  RPCI-11-232E18.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-232E18
            , DNA sequence.
ACCESSION   AQ479538
VERSION     AQ479538.1  GI:4661657
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 182)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1999)
COMMENT     other_GSSs: RPCI-11 232E18.TV
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbestiar.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (bacpac@chori.org), or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.ncbi.nlm.nih.gov/genome/seq/HsBlast.html.
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..182
                     /organism="Homo sapiens"
                     /db_xref="GI:7588817"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-232E18"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI-11 Human Male BAC Library"
BASE COUNT        60 a      32 c      35 g      55 t
ORIGIN


alignment_scores:
            Quality:    6.00            Length:      6
              Ratio:   1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_130_138 x AQ479538    .

 Align seq 1/1  to: AQ479538  from: 1  to: 182

       1 ValLeuLysAspIleGlu 6
         ||||||||||||||||||
      70 CTCCTTAACGACATAGAA 87

Seq_name: qb_qss37:1A152A10F

seq_documentation_block:
LOCUS       1A15-2A10F      202 bp    DNA             GSS       13 DEC 2000
DEFINITION  T. brucei sheared genomic DNA clone 152a10, forward sequence,
            genomic survey sequence.
ACCESSION   AL467175
VERSION     AL467175.1  GI:11837079
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei.
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 202)
  AUTHORS   Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
            Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Bou,I.,
            Melville,S.E., Rajandream,M.A. and Barrell,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (10 DEC 2000) Trypanosoma brucei genome sequencing
            project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
            Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
            nhl@sanger.ac.uk
COMMENT     Constructed at the Institute for Genomic Research (TIGR),
            Rockville, MD. Genomic DNA isolated from a cloned population of
            trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
```

```
            to give a tight size distribution (
            4 kb).  The v e i method used for the library construction is
            described in detail in Smith, H. and Venter, J.C. (Making small
            insert libraries for whole genome shotgun sequencing projects.  In
            Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
            Barrell, Oxford University Press, 1999).
            Email: nelsayed@tigr.org
            Details of T. brucei sequencing at the Sanger Centre are available
            at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES             Location/Qualifiers
     source          1..202
                     /organism "Trypanosoma brucei"
                     /strain "TREU927"
                     /db_xref "taxon:5691"
                     /clone "152a10"
BASE COUNT       91 a      32 c      28 g      51 t
ORIGIN


alignment scores:
             Quality:     6.00              Length:       6
               Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2-COPY_130_138 x TA152A10P  ..

 Align seq 1/1  to: TA152A10P  from: 1  to: 202

       2 LeuLysAspIleGluThr 7
         |||||||||||||||||
      12 CTTAAAGATATAGAAACA 29
```

```
OM of: US-09-126-945B-2_COPY_146_154 to: GenEmbl:*   out_format : pts

Date: Sep 14, 2001  9:47 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL frame+_p2n.model -DEV xlp
 -Q "/cgi-2/tmp/s_spool_us09126945_2.mat 1:%, 1:%,1:%,1:%" -DB_opt_query 1:%,1:%,1:%,1:%
 -DB GenEmbl -QFMT fastap -SUFFIX=olip2n.rge -GAPOP 4.500
 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL 0.000 -LOOPEXT 0.000
 -QGAPOP=4.500 -QGAPEXT 0.050 -XGAPOP 60.000 -XGAPEXT 60.000
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP 60.000 -YGAPEXT 60.000
 -DELOP 6.000 -DELEXT=7.000 -START=1 -MATRIX oligo
 -TRANS human40.cdi -LIST=45 -DOCALIGN 200 -THR_SCORE quality
 -THR_MIN 1  ALIGN 15  MODE LOCAL -OUTFMT pts  NORM=ext  MINLEN 0
 -MAXLEN=2000000000 -USER=US09126945_2CGNI_17594 -NCPU 6
 -ICPU 3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-126-945B-2_COPY_146_154
Query length: 9
Database: GenEmbl:*
Database sequences: 1444157
Database length: -856060004
Search time (sec): 6503.700000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000


score list:
Sequence              Strd Orig    %Score   %Score Log   Length  Documentation
qb_pat1:AX048148       +     9.00  183.22  0.0572     1894  ! AX048148 Sequence 1 from Patent
qb_pr4:AF071538        +     9.00  183.22  0.0572     1894  ! AF071538 Homo sapiens Ets tran
qb_pr1:AB031549        +     9.00  183.21  0.0572     1898  ! AB031549 Homo sapiens PBP mRNA
qb_pat1:AX048159       +     9.00  180.10  0.0856     3317  ! AX048159 Sequence 12 from Pat
qb_pr6:AL157472              9.00  158.85  1.30     150666  ! AL157472 Human DNA sequence
qb_om:AF342725         +     7.00  149.21  4.18       289  ! AF342725 Bos taurus matrix met
qb_htq21:AL356306      -     7.00  117.90  248.54  107714  ! AL356306 Homo sapiens chrom
qb_pr6:AL139415        -     7.00  116.38  301.69  141428  ! AL139415 Human DNA sequence
qb_htq11:AC026279      -     7.00  115.79  325.50  157159  ! AC026279 Homo sapiens chrom
qb_htq9:AC023591       -     7.00  115.41  346.32  171413  ! AC023591 Homo sapiens chrom
qb_htq19:AC090976      +     7.00  114.95  362.77  182920  ! AC090976 Bos taurus clone RP
qb_pat1:A36745         -     6.00  129.87  53.53      268  ! A36745 Sequence 1 from Patent
qb_pat1:A36746         -     6.00  129.87  53.53      268  ! A36746 Sequence 2 from Patent
qb_pat1:A36747         -     6.00  129.87  53.53      268  ! A36747 Sequence 3 from Patent
qb_pat1:AR066278       -     6.00  129.87  53.53      268  ! AR066278 Sequence 1 from patent
qb_pat1:AR066279       -     6.00  129.87  53.53      268  ! AR066279 Sequence 2 from patent
qb_pat1:AR066280       -     6.00  129.87  53.53      268  ! AR066280 Sequence 3 from patent
qb_pat2:I33087               6.00  129.87  53.53      268  ! I33087 Sequence 1 from patent U
qb_pat2:I33088               6.00  129.87  53.53      268  ! I33088 Sequence 2 from patent U
qb_pat2:I33089         -     6.00  129.87  53.53      268  ! I33089 Sequence 3 from patent U
qb_ba3:UB3G16SRR       -     6.00  129.28  57.74      298  ! Y12796 Unidentified bacterium
qb_pat1:AR066300       -     6.00  128.99  59.94      314  ! AR066300 Sequence 23 from paten
qb_pat2:I33109         -     6.00  128.99  59.94      314  ! I33109 Sequence 23 from patent
qb_pat1:A30396         -     6.00  128.51  63.71      342  ! A30396 M.tortuitum DNA for 65 k
qb_pat1:A30397         +     6.00  128.51  63.71      342  ! A30397 M.tortuitum DNA for 65 k
qb_pat1:A30752         -     6.00  128.51  63.71      342  ! A30752 Mycobacterial sp. DNA fo
qb_pat1:A30753         +     6.00  128.51  63.71      342  ! A30753 Mycobacterial sp. DNA fo
qb_ba2:AF257467        -     6.00  128.50  63.84      343  ! AF257467 Mycobacterium peregri
qb_ba2:AF257468        -     6.00  128.50  63.84      343  ! AF257468 Mycobacterium terrae
qb_ba3:S76645          -     6.00  128.50  63.84      343  ! S76645 65 kda antigen [Mycobact
qb_ba3:S76645          -     6.00  128.50  63.84      343  ! S76645 65 kda antigen [Mycobact
qb_ba3:S76648          -     6.00  128.50  63.84      343  ! S76648 65 kda antigen [Mycobact
qb_pat1:AR066299       -     6.00  128.38  64.77      350  ! AR066299 Sequence 22 from paten
qb_pat2:I33108         -     6.00  128.38  64.77      350  ! I33108 Sequence 22 from patent
qb_ba2:AF241203        -     6.00  128.23  66.08      360  ! AF241203 Mycobacterium avium c
qb_ba3:MBU17925        -     6.00  128.23  66.08      360  ! U17925 Mycobacterium bovis 65 k
qb_ba3:MCU17926              6.00  128.23  66.08      360  ! U17926 Mycobacterium chelonae s
qb_ba3:MCU17927              6.00  128.23  66.08      360  ! U17927 Mycobacterium chelonae s
qb_ba3:MCU17928        -     6.00  128.23  66.08      360  ! U17928 Mycobacterium chelonae s
qb_ba3:MPU17952              6.00  128.23  66.08      360  ! U17952 Mycobacterium phlei 6
qb_ba3:MTU17957              6.00  128.23  66.08      360  ! U17957 Mycobacterium tuberculo
qb_ba3:MTMEWPA               6.00  128.19  66.24      665  ! X81862 M.tuberculosis (patie
qb_ba3:MTMAZWPA        -     6.00  128.12  67.00      367  ! X81861 M.tuberculosis (patie
qb_ba3:MTMPEWPA              6.00  128.06  67.52      371  ! X81863 M.tuberculosis (patie
qb_ba3:MAG310217             6.00  127.91  68.81      381  ! AJ310217 Mycobacterium agri

seq_name: qb_pat1:AX048148

seq_documentation_block:
LOCUS       AX048148     1894 bp    DNA              PAT       15-DEC-2000
DEFINITION  Sequence 1 from Patent WO0070092.
ACCESSION   AX048148
VERSION     AX048148.1  GI:11876967
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1894)
  AUTHORS   Kaufmann,J. and Xin,H.
  TITLE     Expression of ets domain proteins in cancer
  JOURNAL   Patent: WO 0070092-A 1 23-NOV-2000;
            Chiron Corporation (US)
FEATURES             Location/Qualifiers
     source          1..1894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             416..1423
                     /note="human JEN1S."
                     /codon_start=1
                     /protein_id="CAC19002.1"
                     /db_xref="GI:11876968"
                     /translation="MSSASPSLSSVSSPSHLLLPPSVSPSTLEKAAASAVSLERRPWS
                     PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAPGASSREEPPPEPEQCPVTDSCAPA
                     GSLRLVPGGLTLEEHSLEQVQSMVVGEVLKDHELACKLLNITALPMEWSFSNVQKWLL
                     WTFEQYELPFMSFAFQSEAGFELCAMSEFQFPQPSPLSGDVEHAEELGWFSAAWMKER
                     LSPCATRSYCASLSEESWLDSEVDSSCSQQPTHLWQFLFELLKKPHSYGRSFEWLNEE
                     GIFELDSAQVAPLWGLRKNPFAMNYDFLSRSIPQIYYFEGLLRKPDISQKLVYQFVHP
                     I"
BASE COUNT       368 a    653 c    571 g    302 t
ORIGIN


alignment scores:
            Quality:    9.00              Length:      9
              Ratio:   1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_146_154 x AX048148  ..

 Align seq 1/1  to AX048148  from: 1  to: 1894

       1 AspProMetAspTrpSerProSerAsn 9
         |||| |||||||||||||||||||||||
     851 GATCCCATGGACTGGAGCCCCAGCAAT 877

seq_name: qb_pr4:AF071538

seq_documentation_block:
LOCUS       AF071538     1894 bp    mRNA             PRI       10-JAN-2000
DEFINITION  Homo sapiens Ets transcription factor PDEF (PDEF) mRNA, complete
            cds.
ACCESSION   AF071538
VERSION     AF071538.1  GI:4007417
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1894)
  AUTHORS   Oettgen,P., Finger,E., Sun,Z., Akbarali,Y., Thamrongsak,U.,
```

```
            Boltax,J., Grall,F., Dube,A., Weiss,A., Brown,L., Quinn,G., Kas,K.,
            Endress,G., Kunsch,C. and Libermann,T.A.
  TITLE     PDEF, a novel prostate epithelium specific Ets transcriptional
            factor, interacts with the androgen receptor and activates
            prostate specific antigen gene expression
  JOURNAL   J. Biol. Chem. 275 (2), 1216-1225 (2000)
  MEDLINE   20092894
REFERENCE   2  (bases 1 to 1894)
  AUTHORS   Libermann,T.A., Oettgen,P., Kunsch,C. and Finger,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUN-1998) Medicine, Beth Israel Deaconess Medical
            Center, 330 Brookline Ave., Boston, MA 02215, USA
FEATURES             Location/Qualifiers
     source          1..1894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6p21.3"
                     /tissue_type="prostate"
     gene            1..1894
                     /gene="PDEF"
     CDS             416..1423
                     /gene="PDEF"
                     /note="prostate epithelium specific"
                     /codon_start=1
                     /product="Ets transcription factor PDEF"
                     /protein_id="AAC95296.1"
                     /db_xref="GI:4007418"
                     /translation="MGSASPGLSSVSPSHLLLPPDIVSKTGLEKAAAGAVGLERRDWS
                     PSPPATPEQGLSAEYLSYFDMLYPEDSSWAAKAPGASSKEEPPEEPEQCPVIDSQAPA
                     GSLDLVPGGLTLEEHSLEQVQSMVVGEVLKDIETACKLLNLTADPMDWSPSNVQKWLL
                     WTEHQYRLPPMGKAFQELAGKELCAMSEEQFEQRSPLGGDVLHAHLDIWKSAAWMKER
                     TSPGATHYCASISEESWIDSEVDSSCSGQFTHLWQFLKELLLKPHSYGKFTKWLNKEK
                     GIFKIEDSAQVARLWGIRKNKPAMNYDKLSRSLRQYYKKGIIERKPDISQRLVYQFVHP
                     I"
BASE COUNT       468 a    653 c    571 g    302 t
ORIGIN


alignment scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY-146-154 x AF071548   ..

 Align seq 1/1  to: AF071548  from: 1  to: 1894

        1 AspProMetAspTrpSerProSerAsn 9
          |||||||||||||||||||||||||||
      851 GATCCCATGGACTGGAGCCCCAGCAAT 877

seq_name: qb_pri:AB031549

seq_documentation_block:
LOCUS       AB031549      1898 bp    mRNA            PRI       20-JAN-2000
DEFINITION  Homo sapiens PSE mRNA for prostate ets, complete cds.
ACCESSION   AB031549
VERSION     AB031549.1  GI:6721497
KEYWORDS    prostate ets.
SOURCE      Homo sapiens adult male prostate epithelium prostate cancer
            cell line:PC-3 cDNA to mRNA
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Yamada,N., Tamai,Y., Miyamoto,H. and Nozaki,M.
  TITLE     Cloning and expression of the mouse Pse gene encoding a novel Ets
            family member
  JOURNAL   Gene 241 (2), 267-274 (2000)
  MEDLINE   20137504
REFERENCE   2  (bases 1 to 1898)
  AUTHORS   Nozaki,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUN-1998) To the DDBJ/EMBL/GenBank databases. Masami
            Nozaki, Osaka University, Institute for Microbial Diseases; 3-1,
            Yamadaoka, Suita, Osaka 565-0871, Japan
            (E-mail:mnozaki@biken.osaka-u.ac.jp, Tel:81-6-879-8348,
            Fax:81-6-879-8349)
FEATURES             Location/Qualifiers
     source          1..1898
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="PC-3"
                     /cell_type="prostate cancer"
                     /chromosome="6"
                     /dev_stage="adult"
                     /map="6p21.2-21.3"
                     /sex="male"
                     /tissue_type="prostate epithelium"
                     /note="vector:pUC18"
     gene            406..1413
                     /gene="PSE"
     CDS             406..1413
                     /gene="PSE"
                     /codon_start=1
                     /product="prostate ets"
                     /protein_id="BAA89544.1"
                     /db_xref="GI:6721498"
                     /translation="MGSASPGLSSVSPSHLLLPPDIVSKTGLEKAAAGAVGLERRDWS
                     PSPPATPEQGLSAEYLSYFDMLYPEDSSWAAKAPGASSKEEPPEEPEQCPVIDSQAPA
                     GSLDLVPGGLTLEEHSLEQVQSMVVGEVLKDIETACKLLNLTADPMDWSPSNVQKWLL
                     WTEHQYRLPPMGKAFQELAGKELCAMSEEQFEQRSPLGGDVLHAHLDIWKSAAWMKER
                     TSPGATHYCASISEESWIDSEVDSSCSGQFTHLWQFLKELLLKPHSYGKFTKWLNKEK
                     GIFKIEDSAQVARLWGIRKNKPAMNYDKLSRSLRQYYKKGIIERKPDISQRLVYQFVHP
                     I"
BASE COUNT       980 a    651 c    567 g    300 t
ORIGIN


alignment scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY-146-154 x AB031549   ..

 Align seq 1/1  to: AB031549  from: 1  to: 1898

        1 AspProMetAspTrpSerProSerAsn 9
          |||:|||||||||||||||||||||||||
      841 GATCCCATGGACTGGAGCCCCAGCAAT 867

seq_name: qb_pat1:AX048159

seq_documentation_block:
LOCUS       AX048159      3317 bp    DNA             PAT       15-DEC-2000
DEFINITION  Sequence 12 from Patent WO0070092.
ACCESSION   AX048159
VERSION     AX048159.1  GI:11876975
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3317)
  AUTHORS   Kaufmann,J. and Xin,H.
  TITLE     Expression of ets domain proteins in cancer
  JOURNAL   Patent: WO 0070092-A 12 23-NOV-2000;
            Chiron Corporation (US)
FEATURES             Location/Qualifiers
     source          1..3317
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

```
                     /note "GSEF encoding sequence with promoter (Frds. 2A 2B)"
     promoter        1. .1480
                     /note "5' to "5' ..."
     CDS             1815. .2822
                     /note "GSEF coding sequence"
                     /codon_start 1
                     /protein_id "CAC19004.1"
                     /db_xref "GI:11876976"
                     /translation "MGSASSQLPSVSFSFRLLLPSSVVSSSFFFAAACAVCFFFWC
                     PSPFATPFSGLSAFYLSFFLMLYPFLSSWAASAFCASSFFFFFFFSSSFDVLLSSAFA
                     GSLDLVPGGLILEFHSLFQVQSMVVGEVLKDIEIACKLLNILALPMDWSPSNVQKWLL
                     WTEHQYRLPFMGKAFQFLACKKLCAMSFLFLFSFFCCFVLRAGLSWKSAAWMFFF
                     TSPGAIHYCASISFFSWTDSFVFSSCSSQLFRLWQFLKFLLKFHSYCSFFFWLNFFK
                     GIFKIFDSAQVAPLWSIPPYSFAMSYFFLSSCFFLYYFFYLFFFLFQSLVY_IVSSF
                     I"
BASE COUNT       710 a   1026 c    976 g    611 t
ORIGIN


Alignment_scores:
            Quality:     9.00              Length:       9
              Ratio:    1.000               Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_146_154 x AX048159     .

 Align seq 1/1  to: AX048159  from: 1  to: 3417

       1 AspProMetAspIrpSerProSerAsn 9
         |||||||||||||||||||||||||||
    2250 GATCCCATGGACTGGAGCCCCAGCAAT 2276

seq_name: gb_pr6:AL157372

seq_documentation_block:
LOCUS        AL157372    150660 bp    DNA              PRI       29 NOV 2000
DEFINITION   Human DNA sequence from clone RP11-375E1 on chromosome 6, complete
             sequence.
ACCESSION    AL157372
VERSION      AL157372.18  GI:11493229
KEYWORDS     HTG.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 150660)
  AUTHORS    Skuce,C.
  TITLE      Direct Submission
  JOURNAL    Submitted (29 NOV 2000) Sanger Centre, Hinxton, Cambridgeshire,
             CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
             requests: clonerequest@sanger.ac.uk
COMMENT      On Nov 30, 2000 this sequence version replaced gi:11342772.
             During sequence assembly data is compared from overlapping clones.
             Where differences are found these are annotated as variations
             together with a note of the overlapping clone name. Note that the
             variation annotation may not be found in the sequence submission
             corresponding to the overlapping clone, as we submit sequences with
             only a small overlap as described above.
             This sequence has been finished according to sequence map criteria
             as follows. An attempt is made to resolve all sequencing problems,
             such as compressions and repeats, but not necessarily within known
             annotated repeat sequence elements. Where the sequence is
             ambiguous, there is an annotation using the 'unsure' feature key.
             The following abbreviations are used to associate primary accession
             numbers given in the feature table with their source databases:
             Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP. Information
             on the WORMPEP database can be found at
             http://www.sanger.ac.uk/Projects/C_elegans/wormpep. This sequence
             was generated from part of bacterial clone contigs of human
             chromosome 6, constructed by the Sanger Centre chromosome 6 Mapping
             Group. Further information can be found at
             http://www.sanger.ac.uk/HGP/Chr6.
             RP11-375E1 is from the library RPCI-11.2 constructed at the Roswell
             Park Cancer Institute by the group of Pieter de Jong. For further
             details see http://bacpac.chori.org/library.htm#rpci_11_2
             VECTOR: pBACe3.6
             IMPORTANT: This sequence is not the entire insert of clone
             RP11-375E1. It may be shorter because we sequence overlapping
             sections only once, except for a 100 base overlap.
             The true left end of clone RP11-375E1 is at 1 in this sequence. The
             true left end of clone RP3-302c22 is at 150661 in this sequence.
             The true right end of clone RP11-302L16 is at 117744 in this
             sequence.
FEATURES             Location/Qualifiers
     source          1. .150660
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome "6"
                     /clone "RP11-375E1"
                     /clone_lib="RPCI-11.2"
BASE COUNT    35470 a  38887 c  38740 g  37563 t
ORIGIN


alignment_scores:
            Quality:     9.00              Length:       9
              Ratio:    1.000               Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_146_154 x AL157372/rev  ..

 Align seq 1/1  to reverse of: AL157372  from: 1  to: 150660

       1 AspProMetAspIrpSerProSerAsn 9
         |||||||||||||||||||||||||||
  126467 GATCCCATGGACTGGAGCCCCAGCAAT 126441

seq_name: gb_om:AF342725

seq_documentation_block:
LOCUS        AF342725       389 bp    DNA              MAM       05 APR 2001
DEFINITION   Bos taurus matrix metalloproteinase 1 (MMP1) gene, partial
             sequence.
ACCESSION    AF342725
VERSION      AF342725.1  GI:13549065
KEYWORDS     .
SOURCE       cow.
  ORGANISM   Bos taurus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
             Bovinae; Bos.
REFERENCE    1  (bases 1 to 389)
  AUTHORS    Christison,K.M., Schmutz,S.M., Berryere,T.G. and Buchanan,F.C.
  TITLE      Linkage Mapping of MMP1 to Cattle Chromosome 15
  JOURNAL    Anim. Genet. (2001) In press
REFERENCE    2  (bases 1 to 389)
  AUTHORS    Christison,K.M. and Schmutz,S.M.
  TITLE      Direct Submission
  JOURNAL    Submitted (25 JAN 2001) Animal Science, University of Saskatchewan,
             51 Campus Dr., Saskatoon, Sask S7N 5A8, Canada
FEATURES             Location/Qualifiers
     source          1. .389
                     /organism="Bos taurus"
                     /db_xref "taxon:9913"
                     /chromosome "15"
     gene            <1. .>389
                     /gene "MMP1"
                     /note "matrix metalloproteinase 1; coding region not
                     determined; similar to GenBank Accession Number X58250"
BASE COUNT       92 a    105 c     94 g     98 t
ORIGIN


alignment_scores:
```

```
            Quality:     7.00              Length:       7
              Ratio:    1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_146_154 x AF342725  ..

 Align seq 1/1  to: AF342725  from: 1  to: 389

        1 AspProMetAspTrpSerPro 7
          |||||||||||||||||||||
      219 GACCCCATGGACTGGAGCCCA 239

Seq_name: qb htq21:AL356306

seq_documentation block:
LOCUS       AL356306   107714 bp    DNA              HTG       23-JAN-2001
DEFINITION  Homo sapiens chromosome 1 clone RP4-684L19 map p36.21-36.33, ***
            SEQUENCING IN PROGRESS ***, 10 unordered pieces.
ACCESSION   AL356306
VERSION     AL356306.4  GI:9863760
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 107714)
  AUTHORS   Pavitt,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@Sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 21, 2000 this sequence version replaced gi:9213829.
                    - Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
                    -- Project Information
            Center project name: dJ684L19
                    -- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: M13; M77815; 2% of reads
            Sequencing vector: plasmid; L08752; 97% of reads
            Chemistry: Dye-terminator ABI; 2% of reads
            Chemistry: Dye terminator BigDye; 97% of reads
            Consensus quality: 102398 bases at least Q40
            Consensus quality: 104160 bases at least Q30
            Consensus quality: 105254 bases at least Q20
            Insert size: 106814; sum of contigs
            Insert size: 125458; 1.4% error; agarose fp
            Quality coverage: 4.87x in Q20 bases; sum of contigs quality
            coverage: 5.39x in Q20 bases; agarose fp
                    -.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    20487: contig of 20487 bp in length
            *    20488 20587: gap of      100 bp
            *    20588    49082: contig of 28495 bp in length
            *    49083 49182: gap of      100 bp
            *    49183    52685: contig of 3503 bp in length
            *    52686 52785: gap of      100 bp
            *    52786    60178: contig of 7393 bp in length
            *    60179 60278: gap of      100 bp
            *    60279    66201: contig of 5923 bp in length
            *    66202 66301: gap of      100 bp
            *    66302    83567: contig of 17266 bp in length
            *    83568 83667: gap of      100 bp
            *    83668    96991: contig of 13324 bp in length
            *    96992 97091: gap of      100 bp
            *    97092    99503: contig of 2412 bp in length
            *    99504 99603: gap of      100 bp
            *    99604   102213: contig of 2610 bp in length
            *   102214 102313: gap of      100 bp
            *   102314   107714: contig of 5401 bp in length.
FEATURES             Location/Qualifiers
     source          1..107714
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="p36.21-36.33"
                     /clone="RP4-684L19"
                     /clone_lib="RPCI-4"
     misc_feature    1..20487
                     /note="assembly_fragment:00387
                     fragment_chain:1"
     misc_feature    20588..49082
                     /note="assembly_fragment:01016
                     fragment_chain:1"
     misc_feature    49183..52685
                     /note="assembly_fragment:00915
                     fragment_chain:1"
     misc_feature    52786..60178
                     /note="assembly_fragment:00820
                     fragment_chain:1"
     misc_feature    60279..66201
                     /note="assembly_fragment:00242
                     fragment_chain:1"
     misc_feature    66402..83567
                     /note="assembly_fragment:00499
                     fragment_chain:2"
     misc_feature    83668..96991
                     /note="assembly_fragment:00468
                     fragment_chain:2"
     misc_feature    97092..99503
                     /note="assembly_fragment:00315
                     fragment_chain:2"
     misc_feature    99604..102213
                     /note="assembly_fragment:00500
                     fragment_chain:2"
     misc_feature    102314..107714
                     /note="assembly_fragment:00991"
BASE COUNT     22992 a  25474 c  26217 g  27027 t    904 others
ORIGIN


alignment scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_146_154 x AL356306/rev  ..

 Align seq 1/1  to reverse of: AL356306  from: 1  to: 107714

       2 ProMetAspTrpSerProSer 8
         ||||||||||||||||||||
   74447 CCCATGGACTGGAGCCCCTCA 74427

Seq_name: qb pri:AL139415

seq_documentation block:
LOCUS       AL139415   141428 bp    DNA              PRI       05-APR-2001
DEFINITION  Human DNA sequence from clone RP3-477M7 on chromosome 1, complete
            sequence.
ACCESSION   AL139415
VERSION     AL139415.10  GI:13559994
KEYWORDS    HTG.
SOURCE      human.
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 141428)
  AUTHORS   Frankland,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: hgmquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 6, 2001 this sequence version replaced gi:13274269.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted, all
            regions were either double stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e. phrap quality >
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats, all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group. Further information can be found at
            http://www.sanger.ac.uk/Chr1
            RP3-477M7 is from the library RPCI-3 constructed by the group of
            Pieter de Jong. For further details see
            http://bacpac.chori.org/bacpac_home.htm
            VECTOR: pCYPAC2
            This sequence is the entire insert of clone RP3-477M7.
FEATURES             Location/Qualifiers
     source          1..141428
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP3-477M7"
                     /clone_lib="RPCI-3"
     repeat_region   294..435
                     /note="AluJo repeat: matches 1..140 of consensus"
     repeat_region   436..747
                     /note="AluSq repeat: matches 1..310 of consensus"
     repeat_region   748..894
                     /note="AluJo repeat: matches 140..297 of consensus"
     repeat_region   994..1052
                     /note="L1ME3 repeat: matches 5987..6047 of consensus"
     repeat_region   1089..1264
                     /note="THE1C repeat: matches 1..144 of consensus"
     repeat_region   1265..1577
                     /note="AluSx repeat: matches 2..311 of consensus"
     repeat_region   1578..1803
                     /note="THE1C repeat: matches 133..371 of consensus"
     repeat_region   1941..2213
                     /note="AluSx repeat: matches 21..293 of consensus"
     repeat_region   2453..2454
                     /note="MIR repeat: matches 80..172 of consensus"
     repeat_region   2794..2958
                     /note="MER5B repeat: matches 1..178 of consensus"
     repeat_region   2979..3115
                     /note="FLAM_A repeat: matches 1..125 of consensus"
     repeat_region   3141..3430
                     /note="AluSx repeat: matches 1..290 of consensus"
     repeat_region   4013..4159
                     /note="HERVL repeat: matches 5615..5764 of consensus"
     repeat_region   4162..4320
                     /note="MLT2E repeat: matches 14..188 of consensus"
     repeat_region   4321..4614
                     /note="AluSx repeat: matches 1..296 of consensus"
     repeat_region   4615..4706
                     /note="MLT2E repeat: matches 188..275 of consensus"
     repeat_region   4790..4944
                     /note="AluJb repeat: matches 130..286 of consensus"
     repeat_region   5140..5299
                     /note="AluSx repeat: matches 136..305 of consensus"
     repeat_region   5300..5595
                     /note="AluSq repeat: matches 1..296 of consensus"
     repeat_region   5596..5720
                     /note="AluSx repeat: matches 12..136 of consensus"
     misc_feature    complement(6425..6778)
                     /note="match: STS: Em:G11243"
     misc_feature    6456..6761
                     /note="match: STS: Em:G05574"
     repeat_region   7323..7602
                     /note="AluSx repeat: matches 31..310 of consensus"
     repeat_region   7605..7700
                     /note="AluJ/FRAM repeat: matches 209..304 of consensus"
     repeat_region   9055..9140
                     /note="FLAM_C repeat: matches 21..106 of consensus"
     repeat_region   9747..10024
                     /note="AluYo repeat: matches 1..278 of consensus"
     repeat_region   10174..10317
                     /note="MIR repeat: matches 7..163 of consensus"
     repeat_region   11412..11523
                     /note="AluSg repeat: matches 193..309 of consensus"
     misc_feature    complement((11718..11720,12520,12536..12669,13402..13428))
                     /note="match: STS: Em:G37215"
     repeat_region   12096..12223
                     /note="FLAM_A repeat: matches 2..121 of consensus"
     repeat_region   12289..12424
                     /note="MIR repeat: matches 62..191 of consensus"
     repeat_region   12972..13248
                     /note="AluY repeat: matches 1..302 of consensus"
     repeat_region   14096..14406
                     /note="AluJb repeat: matches 1..309 of consensus"
     repeat_region   14978..15278
                     /note="AluSx repeat: matches 1..301 of consensus"
     repeat_region   17640..17692
                     /note="MER5A repeat: matches 30..82 of consensus"
     repeat_region   17747..17805
                     /note="Charlie5 repeat: matches 2527..2585 of consensus"
     repeat_region   17807..18099
                     /note="AluSx repeat: matches 3..296 of consensus"
     repeat_region   18584..18750
                     /note="L1ME1 repeat: matches 6148..6316 of consensus"
     repeat_region   19713..20025
                     /note="AluSx repeat: matches 4..311 of consensus"
     repeat_region   20561..20685
                     /note="MER85 repeat: matches 1..136 of consensus"
     repeat_region   20717..20750
                     /note="17 copies 2 mer aa 82% conserved"
     repeat_region   21208..21339
                     /note="FLAM_C repeat: matches 1..124 of consensus"
     repeat_region   21853..21930
                     /note="L1ME1 repeat: matches 5470..5605 of consensus"
     repeat_region   22046..22245
                     /note="AluY repeat: matches 2..302 of consensus"
     repeat_region   22264..22407
                     /note="AluJo repeat: matches 138..299 of consensus"
     repeat_region   22412..22670
                     /note="AluY repeat: matches 1..256 of consensus"
     repeat_region   22678..22810
                     /note="AluJo repeat: matches 3..130 of consensus"
     repeat_region   22993..23310
                     /note="AluSp repeat: matches 1..298 of consensus"
     misc_feature    24150..24769
                     /note="CpG island"
                     /evidence=not_experimental
     repeat_region   26524..26833
                     /note="AluSq repeat: matches 2..310 of consensus"
     repeat_region   26856..27152
```

```
                     /note "AluJo repeat: matches 4, .295 of consensus"
     repeat_region   27156, .27469
                     /note "AluSq repeat: matches 1, .313 of consensus"
     repeat_region   27480, .27780
                     /note "AluSx repeat: matches 1, .302 of consensus"
     repeat_region   27874, .28179
                     /note "AluY repeat: matches 1, .307 of consensus"
     repeat_region   28203, .28505
                     /note "AluSq repeat: matches 1, .298 of consensus"
     repeat_region   28508, .28813
                     /note "AluY repeat: matches 1, .306 of consensus"
     repeat_region   28848, .28944
                     /note "U6 repeat: matches 14, .105 of consensus"
     repeat_region   29398, .29697
                     /note "AluSp repeat: matches 1, .301 of consensus"
     repeat_region   29955, .30261
                     /note "AluSx repeat: matches 1, .306 of consensus"
     repeat_region   30823, .31110
                     /note "AluSq repeat: matches 1, .291 of consensus"
     repeat_region   31450, .31751
                     /note "AluSx repeat: matches 1, .302 of consensus"
     repeat_region   31752, .32010
                     /note "MLT1A1 repeat: matches 76, .349 of consensus"
     repeat_region   32138, .32215
                     /note "MER58A repeat: matches 1, .89 of consensus"
     repeat_region   32216, .32431
                     /note "AluJb repeat: matches 87, .297 of consensus"
     repeat_region   32654, .32704
                     /note "MLT1I repeat: matches 91, .137 of consensus"
     repeat_region   32808, .32822
                     /note "L2 repeat: matches 2074, .2087 of consensus"
     repeat_region   32823, .33130
                     /note "AluSq repeat: matches 1, .308 of consensus"
     repeat_region   33131, .33421
                     /note "AluJo repeat: matches 1, .293 of consensus"
     repeat_region   33422, .33545
                     /note="L2 repeat: matches 1955, .2073 of consensus"
     repeat_region   33550, .33844
                     /note="AluJo repeat: matches 1, .295 of consensus"
     repeat_region   33848, .34140
                     /note "AluSq repeat: matches 1, .297 of consensus"
     repeat_region   34144, .34450
                     /note "AluY repeat: matches 1, .310 of consensus"
     repeat_region   34489, .34792
                     /note "AluJb repeat: matches 5, .306 of consensus"
     repeat_region   35712, .35752
                     /note "Alu repeat: matches 260, .360 of consensus"

alignment scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_146_154 x AL139415/rev   ..

 Align seq 1/1  to reverse of AL139415  from  1  to  141428

        2 ProMetAspTrpSerProSer 8
          ||||||||||||||||||||||
   94899 CCCATGGACTGGAGCCCCTCA 94879

seq name: qb_htqll:AC026279

seq documentation block:
LOCUS       AC026279    157159 bp    DNA              HTG       07-JUL-2000
DEFINITION  Homo sapiens chromosome 15 clone RP11-321G12, WORKING DRAFT
            SEQUENCE, 12 unordered pieces.
ACCESSION   AC026279
VERSION     AC026279.4  GI:8954410
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 157159)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 157159)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jul 7, 2000 this sequence version replaced gi:7524955.

            -- - - Genome Center - -
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc/index.shtml
            - - - Project Information - - -
            Center project name: H_NH0321G12
            - - - - Summary Statistics - - -
            Sequencing vector: M13; 86%
            Sequencing vector: plasmid; 14%
            Chemistry: Dye-primer ET; 86% of reads
            Chemistry: Dye terminator Big Dye; 14% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 148266 bases at least Q40
            Consensus quality: 151490 bases at least Q30
            Consensus quality: 153444 bases at least Q20
            Insert size: 165000; agarose fp
            Insert size: 156059; sum of contigs
            Quality coverage: 4.15 in Q20 bases; agarose fp
            Quality coverage: 3.96 in Q20 bases; sum of contigs
            - - - - - - - - - - - - - - - - - - - - - - - - - -
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      1812: contig of 1812 bp in length
            *     1813      1912: gap of unknown length
            *     1913      4050: contig of 2138 bp in length
            *     4051      4150: gap of unknown length
            *     4151      8892: contig of 4742 bp in length
            *     8893      8992: gap of unknown length
            *     8993     14338: contig of 5346 bp in length
            *    14339     14438: gap of unknown length
            *    14439     20060: contig of 5622 bp in length
            *    20061     20160: gap of unknown length
            *    20161     27244: contig of 7084 bp in length
            *    27245     27344: gap of unknown length
            *    27345     42412: contig of 15068 bp in length
            *    42413     42512: gap of unknown length
            *    42513     55808: contig of 13296 bp in length
            *    55809     55908: gap of unknown length
            *    55909     69598: contig of 13690 bp in length
            *    69599     69698: gap of unknown length
            *    69699     88164: contig of 18466 bp in length
            *    88165     88264: gap of unknown length
            *    88265    106938: contig of 18674 bp in length
            *   106939    107038: gap of unknown length
            *   107039    157159: contig of 50121 bp in length.
FEATURES             Location/Qualifiers
     source          1..157159
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /clone="RP11-321G12"
     misc_feature    1..1812
                     /note "assembly_name:Contig1"
```

```
     misc_feature    1913. .4050
                     /note="assembly_name:Contig2"
     misc_feature    4151. .8892
                     /note="assembly_name:Contig3"
     misc_feature    8993. .14338
                     /note="assembly_name:Contig4"
     misc_feature    14439. .20060
                     /note="assembly_name:Contig5"
     misc_feature    20161. .27244
                     /note="assembly_name:Contig6"
     misc_feature    27345. .42412
                     /note="assembly_name:Contig7"
     misc_feature    42513. .55808
                     /note="assembly_name:Contig8"
     misc_feature    55909. .69598
                     /note="assembly_name:Contig9"
     misc_feature    69699. .88164
                     /note="assembly_name:Contig10"
     misc_feature    88265. .106938
                     /note="assembly_name:Contig11
                     clone_end:SP6
                     vector_side:left"
     misc_feature    107039. .157159
                     /note="assembly_name:Contig12
                     clone_end:T7
                     vector_side:left"
BASE COUNT    42295 a   34576 c   35453 g   43728 t    1107 others
ORIGIN


alignment_scores:
            Quality:    7.00            Length:       7
              Ratio:  1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_146_154 x AC026279/rev  ..

 Align seq 1/1  to reverse of. AC026279  from. 1  to. 157159

       2 ProMetAspTrpSerProSer 8
         ||||||||||||||||||||
   17439 CCCATGGACTGGAGTCCCTCC 17419

seq_name: qb_htq9:AC023591

seq_documentation_block:
LOCUS        AC023591    171413 bp     DNA              HTG       16-MAR-2001
DEFINITION   Homo sapiens chromosome 11 clone RP11-768L13 map 11, WORKING DRAFT
             SEQUENCE, 5 unordered pieces.
ACCESSION    AC023591
VERSION      AC023591.3  GI:13357465
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 171413)
  AUTHORS    Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE      Homo sapiens chromosome 11, clone RP11-768L13
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 171413)
  AUTHORS    Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
             Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslawsky,I.,
             Boukhalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
             Choepel,Y., Colanqelo,M., Collins,S., Collymore,A., Cooke,P.,
             DeArellano,K., Dewar,K., Dodqe,S., Domino,M., Doyle,M.,
             Fenestor,J., Ferreira,P., FitzHuqh,W., Forrest,C., Gaqe,D.,
             Galaqan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
             Grand Pierre,N., Grant,G., Haqos,B., Heaford,A., Horton,L.,
             Howland,J.C., Iliev,I., Johnson,E., Jones,C., Kann,L., Karatas,A.,
             Klein,T., Linders,T., Largespe,K., Lebovsky,I., Levine,R.,
             Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
             McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
             Menzes,S., Milosa,I., Miranda,C., Mlenqa,V., Morrow,J., Nayor,J.,
             Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,I.M.,
             Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
             Riley,R., Roqov,P., Rothman,D., Roy,A., Santos,R., Schaper,S.,
             Sewery,P., Spencer,B., Stampe Thomann,N., Stojanovic,N.,
             Subramanian,A., Talamas,J., Testayo,S., Theodore,J., Tirrell,A.,
             Travers,M., Trifillio,J., Vassiliev,H., Viel,R., Vo,A., Wils on,K.,
             Wu,X., Wyman,D., Yo,W.J., Young,G., Zaitoun,J., Zimmer,A. and
             Zody,M.
  TITLE      Direct Submission
  JOURNAL    Submitted (27-FEB-2001) Whitehead Institute/MIT Center for Genome
             Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT      On Mar 16, 2001 this sequence version replaced gi:7210084.
             All repeats were identified using RepeatMasker:
             Smit, A.F.A. & Green, P. (1996-1997)
             http://ftp.genome.washington.edu/RM/RepeatMasker.html
              ---------- - Genome Center
                 Center: Whitehead Institute/ MIT Center for Genome Research
                 Center code: WIBR
                 Web site: http://www-seq.wi.mit.edu
                 Contact: sequence_submissions@genome.wi.mit.edu
               --- ---   - Project Information
                 Center project name: L6849
                 Center clone name: 768_L_13
               ---- ---   - Summary Statistics
                 Sequencing vector: M13; M77815; 36% of reads
                 Sequencing vector: Plasmid; n/a; 64% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.960731
                 Consensus quality: 169848 bases at least Q40
                 Consensus quality: 170365 bases at least Q30
                 Consensus quality: 170681 bases at least Q20
                 Insert size: 165000; agarose-fp
                 Insert size: 171013; sum-of-contigs
                 Quality coverage: 10.5 in Q20 bases; agarose-fp
                 Quality coverage: 10.2 in Q20.
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 5 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence
             * as soon as it is available and the accession number will
             * be preserved.
             *       1    20222: contig of 20222 bp in length
             *   20223 20322: gap of      100 bp
             *   20323    21049: contig of 727 bp in length
             *   21050 21149: gap of      100 bp
             *   21150    89654: contig of 68505 bp in length
             *   89655 89754: gap of      100 bp
             *   89755   153697: contig of 63943 bp in length
             * 153698 153797: gap of      100 bp
             * 153798   171413: contig of 17616 bp in length
FEATURES             Location/Qualifiers
     source          1. .171413
                     /organism="Homo sapiens"
                     /db_xref "taxon:9606"
                     /chromosome "11"
                     /map "11"
                     /clone "RP11-768L13"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .20222
                     /note "assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    20323. .21049
                     /note="assembly_fragment"
     misc_feature    21150. .89654
                     /note="assembly_fragment"
     misc_feature    89755. .153697
                     /note "assembly_fragment"
     misc_feature    153798. .171413
```

```
                     /note "assembly fragment
                     clone end:I7
                     vector_side:right"
BASE COUNT    45485 a   39846 c   39175 g   46507 t     400 others
ORIGIN


alignment scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_146_154 x AC023591/rev   ..

 Align seq 1/1  to reverse of: AC023591  from: 1  to: 171414

       2 ProMetAspIrpSerProSer 8
         |||||||||||||||||||||
   26452 CCCATGGACTGGACTCCCTCC 26432

seq name: gb htg19:AC090976

seq documentation block:
LOCUS       AC090976    182920 bp    DNA             HTG       22-MAR-2001
DEFINITION  Bos taurus clone RP42-400M23, WORKING DRAFT SEQUENCE, 6 unordered
            pieces.
ACCESSION   AC090976
VERSION     AC090976.1  GI:13430990
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos
REFERENCE   1  (bases 1 to 182920)
  AUTHORS   Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,R., Blakesley,R.W.,
            Bouffard,G.G., Brinkley,C., Brooks,S., Dietrich,N.L., Granite,S.,
            Guan,X., Gupta,J., Ho,S.-L., Idol,J.R., Karlins,E., Loftin,S.-Q.,
            Lopaspi,R., Lim,M., Maduro,Q.L., Maduro,V.B., Masiello,C.,
            Mastrian,S.D., McCloskey,J.C., Mebowell,J., Pearson,R., Prasad,A.,
            Shevchenko,Y., Snyder,B., Stantripop,S., Thomas,J.W., Thomas,P.J.,
            Tiongson,E.E., Touchman,J.W., Tsurgeon,C., Voqt,J.L., Walker,M.A.,
            Wetherby,K.D., Zhang,L. B. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 182920)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-2001) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
                --- Genome Center
                Center: NIH Intramural Sequencing Center
                Center code: NISC
                Web site: http://www.nisc.nih.gov
                Contact: nisc_mouse@nhgri.nih.gov
            - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
                -- Project Information
                Center project name: axo
                Center clone name: 400M23
            - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
                -- Summary Statistics
                Sequencing vector: plasmid; n/a; 100% of reads
                Chemistry: dye terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 179499 bases at least Q40
                Consensus quality: 180483 bases at least Q30
                Consensus quality: 181043 bases at least Q20
                Insert size: 141000; agarose tp
                Insert size: 182420; sum of contigs
                Quality coverage: 8.40x in Q30 bases, agarose tp
                Quality coverage: 7.95x in Q20 bases, sum of contigs
            - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     12486: contig of 12486 bp in length
            *    12487     12586: gap of unknown length
            *    12587     25833: contig of 13247 bp in length
            *    25834     25933: gap of unknown length
            *    25934     48044: contig of 22111 bp in length
            *    48045     48144: gap of unknown length
            *    48145     76139: contig of 27995 bp in length
            *    76140     76239: gap of unknown length
            *    76240    118199: contig of 41960 bp in length
            *   118200    118299: gap of unknown length
            *   118300    182920: contig of 64621 bp in length.
FEATURES             Location/Qualifiers
     source          1..182920
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone="RP42-400M23"
                     /clone_lib="RP42"
     misc_feature    1..12486
                     /note="assembly_fragment"
     misc_feature    12587..25833
                     /note="assembly fragment"
     misc_feature    25934..48044
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    48145..76139
                     /note="assembly fragment"
     misc_feature    76240..118199
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    118300..182920
                     /note="assembly_fragment"
BASE COUNT    57077 a   45661 c   45052 g   54607 t     523 others
ORIGIN


alignment scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_146_154 x AC090976   ..

 Align seq 1/1  to: AC090976  from: 1  to: 182920

       1 AspProMetAspIrpSerPro 7
         |||||||||||||||||||||
  141622 GACCCCATGGACTGGAGCCCA 141642

seq name: gb pat1:A36735

seq documentation block:
LOCUS       A36735          268 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 1 from Patent EP0584023.
ACCESSION   A36735
VERSION     A36735.1  GI:2294002
KEYWORDS    .
SOURCE      Mycobacterium bovis.
  ORGANISM  Mycobacterium bovis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE   1  (bases 1 to 268)
  AUTHORS   Matsilat,C. and Pechere,J.
  TITLE     Mycobacteria DNA fragments, amplification primers, hybridization
```

```
            probes, reagents and detection process of mycobacteria
  JOURNAL   Patent: EP 0584023-A 1 23-FEB-1994;
            BIO MERIEUX (FR)
COMMENT     Other publication CA 2103933 940214
            Other publication FR 2694754 940218.
FEATURES             Location/Qualifiers
     source          1. .268
                     /organism="Mycobacterium bovis"
                     /db_xref "taxon:1765"
BASE COUNT       57 a      84 c      94 g      33 t
ORIGIN


alignment_scores:
            Quality:     6.00            Length:       6
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_146_154 x A36735/rev    .

 Align seq 1/1  to reverse of: A36735  from: 1  to: 268

       2 ProMetAspIrpSerPro 7
         |||||||||||||||||||
     261 CCGATGGACTGGTCACCC 244

seq_name: qb_pat1:A36736

seq_documentation_block:
LOCUS       A36736          268 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 2 from Patent EP0584023.
ACCESSION   A36736
VERSION     A36736.1  GI:2294003
KEYWORDS    .
SOURCE      Mycobacterium microti.
  ORGANISM  Mycobacterium microti
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE   1  (bases 1 to 268)
  AUTHORS   Mabilat,C. and Pechere,J.
  TITLE     Mycobacteria DNA fragments, amplification primers, hybridization
            probes, reagents and detection process of mycobacteria
  JOURNAL   Patent: EP 0584023-A 2 23-FEB-1994;
            BIO MERIEUX (FR)
COMMENT     Other publication CA 2103933 940214
            Other publication FR 2694754 940218.
FEATURES             Location/Qualifiers
     source          1. .268
                     /organism="Mycobacterium microti"
                     /db_xref="taxon:1806"
BASE COUNT       57 a      84 c      94 g      33 t
ORIGIN


alignment_scores:
            Quality:     6.00            Length:       6
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_146_154 x A36736/rev    .

 Align seq 1/1  to reverse of: A36736  from: 1  to: 268

       2 ProMetAspIrpSerPro 7
         |||||||||||||||||||
     261 CCGATGGACTGGTCACCC 244

seq_name: qb_pat1:A36737

seq_documentation_block:
LOCUS       A36737          268 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 3 from Patent EP0584023.
ACCESSION   A36737
VERSION     A36737.1  GI:2294004
KEYWORDS    .
SOURCE      Mycobacterium africanum.
  ORGANISM  Mycobacterium africanum
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE   1  (bases 1 to 268)
  AUTHORS   Mabilat,C. and Pechere,J.
  TITLE     Mycobacteria DNA fragments, amplification primers, hybridization
            probes, reagents and detection process of mycobacteria
  JOURNAL   Patent: EP 0584023-A 3 23-FEB-1994;
            BIO MERIEUX (FR)
COMMENT     Other publication CA 2103933 940214
            Other publication FR 2694754 940218.
FEATURES             Location/Qualifiers
     source          1. .268
                     /organism "Mycobacterium africanum"
                     /db_xref "taxon:33894"
BASE COUNT       58 a      84 c      93 g      33 t
ORIGIN


alignment_scores:
            Quality:     6.00            Length:       6
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_146_154 x A36737/rev    .

 Align seq 1/1  to reverse of: A36737  from: 1  to: 268

       2 ProMetAspIrpSerPro 7
         |||||||||||||||||||
     261 CCGATGGACTGGTCACCC 244

seq_name: qb_pat1:AR066278

seq_documentation_block:
LOCUS       AR066278        268 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 1 from patent US 5849901.
ACCESSION   AR066278
VERSION     AR066278.1  GI:5996494
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 268)
  AUTHORS   Mabilat,C. and Pechere,J.
  TITLE     DNA fragments of mycobacteria, amplification primers hybridization
            probes, reagents and method for the detection of mycobacteria
  JOURNAL   Patent: US 5849901-A 1 15-DEC-1998;
FEATURES             Location/Qualifiers
     source          1. .268
                     /organism "unknown"
BASE COUNT       57 a      84 c      94 g      33 t
ORIGIN


alignment_scores:
            Quality:     6.00            Length:       6
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945b-2_COPY_146_154 x AR066278/rev    .

 Align seq 1/1  to reverse of: AR066278  from: 1  to: 268
```

2 ProMetAspErgSerPro 7
|||||||||||||||||||
261 CCGAPGGACTGGTCACCC 244

OM ol: US-09-126-945B-2_COPY_146_154 [illegible]

Date: Sep 11, 2001 [illegible] PM

About: Results were produced by the GenCore software [illegible]
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL frame+_p2n.model -DEV xlp
-Q [illegible]
-DB N_Geneseq_0601 -QFMT=fastap -SUFFIX=olip2n.rng -GAPOP=4.500
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality
-THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -MINLEN=0
-MAXLEN=2000000000 -USER=US09126945_@CGN1_1_308 -NCPU=6 -ICPU=3
-LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
query: US-09-126-945B-2_COPY_146_154
query length: 9
Database: N_Geneseq_0601:*
Database sequences: 730101
Database length: 313950809
Search time (sec): 680.150000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	! Documentation
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF21828 +	9.00	171.54	0.0044	1087	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF50690 +	9.00	170.74	0.1570	1094	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:AAC84261 +	9.00	170.74	0.1570	1094	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:AAC84266 +	9.00	166.92	0.2517	3217	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT67546 +	6.00	125.22	52.76	267	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1994.DAT:AAQ57761 +	6.00	125.15	52.91	268	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1994.DAT:AAQ57762 +	6.00	125.15	52.91	268	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1994.DAT:AAQ57763 +	6.00	125.15	52.91	268	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1990.DAT:AAQ[illegible] +	6.00	123.52	65.54	342	!
/SIDS1/gcgdata/geneseq/geneseqn/NA[illegible].DAT:AAQ[illegible] +	6.00	123.52	65.59	342	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT67868 +	6.00	121.92	82.54	472	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT67864 +	6.00	118.80	129.16	601	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF25022 +	6.00	116.98	151.72	898	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC[illegible] +	6.00	116.78	155.77	515	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1995.DAT:AAT[illegible] +	6.00	114.37	220.54	1359	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF25[illegible] +	6.00	112.05	257.79	1023	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF25[illegible] +	6.00	111.70	[illegible]	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF25019 +	6.00	111.60	392.52	1947	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF25014 +	6.00	109.90	422.14	2917	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV55742 -	6.00	108.17	469.82	4213	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1988.DAT:AAN81760 +	6.00	106.24	[illegible]	1260	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1988.DAT:AAN80222 +	6.00	106.05	[illegible]	1380	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV[illegible] +	6.00	106.05	[illegible]	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX[illegible] +	6.00	105.91	[illegible]	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA[illegible] +	6.00	105.91	[illegible]	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AA[illegible] +	6.00	102.26	690.[illegible]	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA75001 +	5.00	124.68	61.26	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA12700 +	5.00	121.88	91.79	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX[illegible] +	5.00	[illegible]	[illegible]	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT:AAX[illegible] +	5.00	115.80	176.56	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX[illegible] +	5.00	115.80	176.56	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT31[illegible] +	5.00	115.47	184.[illegible]	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT[illegible] +	5.00	115.47	184.[illegible]	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT[illegible] +	5.00	114.75	202.19	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA[illegible] +	5.00	114.18	217.[illegible]	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA[illegible].DAT:AAA[illegible] +	5.00	111.[illegible]	[illegible]	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAX12[illegible] +	5.00	110.72	338.92	126	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC0[illegible] +	5.00	107.66	503.[illegible]	197	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1994.DAT:AAQ[illegible] +	5.00	106.11	612.39	247	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ14144 +	5.00	104.77	726.49	[illegible]	
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ1[illegible] +	5.00	104.77	726.49	[illegible]	
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX98536 +	5.00	104.77	726.49	[illegible]	
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA78190 +	5.00	104.46	756.[illegible]	[illegible]	
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC02689 +	5.00	104.2[illegible]	777.[illegible]	[illegible]	
/SIDS1/gcgdata/geneseq/geneseqn/NA1992.DAT:AAQ30888 +	5.00	104.12	789.92	[illegible]	

seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF21828

seq_documentation_block:
ID AAF21828 standard; DNA; 1087 BP.
XX
AC AAF21828;
XX
DT 27-MAR-2001 (first entry)
XX
DE Human breast and ovarian cancer associated antigen gene SEQ ID 215.
XX
KW Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
KW nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
KW antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
KW Addison's disease; allergy; autoimmune haemolytic anaemia;
KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW multiple sclerosis; rheumatoid arthritis; ulceratize colitis;
KW cardiovascular disorder; wound healing; neurological disease; ds.
XX
OS Homo sapiens.
XX
PN WO200055173-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05881.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-[illegible]58.
DR P-PSDB; AAB58925.
XX
PT New human breast and ovarian cancer associated gene sequences and the
PT polypeptides encoded by these genes, useful in the prevention,
PT treatment and diagnosis of cancer, immune disorders, cardiovascular
PT disorders and neurological diseases -
XX
PS Claim 1; Page [illegible]; 1299pp; English.
XX
CC Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
CC associated with breast and ovarian cancer. Included in the invention are
CC sequences AAF[illegible] - AAF[illegible] and AAB59129 which are used in the
CC [illegible] and characterisation of the DNA and protein sequences of the
CC invention. The breast and ovarian cancer associated DNA, protein, agonist
CC or antagonist sequences exhibit cytostatic; immunosuppressive;
CC nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
CC antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
CC antibacterial; antifungal; antiparasitic and cardiant activity. The
CC polynucleotide and protein sequences are used in the diagnosis of cancer,
CC particularly breast and ovarian cancer. The nucleic acid sequences,
CC proteins, agonists and antagonists may also be used in the diagnosis,
CC prevention and treatment of immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; cardiovascular disorders such as
CC myocardial ischaemias; wound healing; neurological diseases such as
CC cerebral [illegible] and epilepsy; and infectious diseases.
XX
SQ Sequence 1087 BP; 2[illegible] A; 247 C; 339 G; 182 T; 4 other;

```
alignment scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2 (COPY 146_154 x AAF21828    .

 Align_seq 1/1  to: AAF21828  from: 1  to: 1087

        1 AspProMetAspTrpSerProSerAsn 9
          |||||||||||||||||||||||||||
      312 GATCCCATGGACTGGAGCCCCAGCAAT 338

Seq name: /cgi-bin/ftp_data/pto_seq_protein_ftp_DNA_to_...AI_AAZ50691

Seq documentation block:
ID   AAZ50691 standard; cDNA; 1894 BP.
XX
AC   AAZ50691;
XX
DT   23-MAY-2000  (first entry)
XX
DE   Human prostate derived Ets factor cDNA.
XX
KW   Prostate Derived Ets factor; PDEF; human; chromosome 6p21.3; cancer;
KW   loss of heterozygosity; chromosomal translocation; linkage analysis;
KW   cytostatic; cardiant; immunosuppressive; cerebroprotective; fungicide;
KW   antibacterial; vulnerary; neuroprotective; antiparkinsonian; nootropic;
KW   anabolic; antiinflammatory; anorectic; hybridisation probe; forensic;
KW   tumour marker; diagnosis; treatment; prostate; cancer; blood coagulation;
KW   autoimmune disorder; haematopoietic; immune/nervous system; stroke;
KW   neoplasm; microbial infection; tissue regeneration; heart attack;
KW   scarring; food additive; preservative; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             416..1423
FT                   /*tag= a
FT                   /product= "Human prostate derived Ets factor"
FT                   /note= "Expressed in prostate epithelium"
XX
PN   WO200006589-A1.
XX
PD   10-FEB-2000.
XX
PF   02-AUG-1999;   99WO-US17470.
XX
PR   31-JUL-1998;   98US-0126945.
XX
PA   (HUMA ) HUMAN GENOME SCI INC.
PA   (BETH ) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI   Libermann TA, Oettgen JP, Kunsch CA, Endress GA, Rosen CA;
XX
DR   WPI; 2000-195255/17.
DR   P-PSDB; AAY45003.
XX
PT   Novel prostate derived polypeptide, polynucleotide useful for
PT   diagnosis, prevention and treatment of prostate cancer, autoimmune
PT   disorders, microbial infections and also as food additive or
PT   preservative
XX
PS   Claim 4; Fig 1; 132pp; English.
XX
CC   The present cDNA sequence encodes a human prostate derived Ets factor
CC   (PDEF), isolated from human prostate carcinoma cDNA library
CC   (ATCC No.203072). PDEF gene expression is highest in tissues involved in
CC   androgen sensitivity, with enriched expression in prostate epithelium
CC   cells. PDEF gene is mapped to the human chromosome 6p21.3
CC   region that is associated with loss of heterozygosity and chromosomal
CC   translocations in various human cancers. PDEF has cytostatic, cardiant,
CC   immunosuppressive, cerebroprotective, fungicide, antibacterial,
CC   vulnerary, neuroprotective, antiparkinsonian, nootropic, anabolic,
CC   antiinflammatory and anorectic activity. PDEF polynucleotides are useful
CC   in linkage analysis as markers, as hybridisation probes for differential
CC   identification of the tissues or cell types and as polymorphic markers
CC   for forensic purposes. PDEF is useful as prostate-specific tumour marker
CC   for the diagnosis and treatment of prostate cancer. PDEF sequences are
CC   useful for treating autoimmune disorders, haematopoietic, blood
CC   coagulation, immune and nervous system disorders, hyperproliferative
CC   disorders like, neoplasms and microbial infections, heart attacks,
CC   stroke, scarring and for tissue regeneration. They are also useful as
CC   food additives or preservatives.
XX
SQ   Sequence 1894 BP; 368 A; 654 C; 571 G; 302 T; 0 other;


alignment scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2 (COPY 146_154 x AAZ50691    .

 Align_seq 1/1  to: AAZ50691  from: 1  to: 1894

        1 AspPro-MetAspTrpSerProSerAsn 9
          |||||||||||||||||||||||||||
      851 GATCCCATGGACTGGAGCCCCAGCAAT 877

Seq_name: /cgi-bin/ftp_data/pto_seq_protein_ftp_DNA2001.DAT:AAC88261

Seq_documentation block:
ID   AAC88261 standard; DNA; 1894 BP.
XX
AC   AAC88261;
XX
DT   16-MAR-2001  (first entry)
XX
DE   Gland-specific Ets transcription factor (GSEF) nucleotide sequence.
XX
KW   transcription factor; gland-specific Ets transcription factor; GSEF;
KW   metastatic potential; cancer; tumour; metastasis; breast; prostate;
KW   leukaemia; lymphoma; sarcoma; melanoma; chromosome 6p21.1 6p21.3; ds.
XX
OS   Homo sapiens.
XX
PN   WO200070092-A1.
XX
PD   23-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US13173.
XX
PR   14-MAY-1999;   99US-0134112.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Kaufmann E, Xin H, Barrowe G;
XX
DR   WPI; 2001-331375/35.
DR   P-PSDB; AAB40628.
XX
PT   Detecting metastatic and potential metastatic cancerous cells, useful
PT   for diagnosing, prognosing, grading and staging of cancers by detecting
PT   gland specific Ets transcription factor gene product in a biological
PT   sample from a cell
XX
PS   Disclosure; Page 81-83; 95pp; English.
XX
CC   This invention relates to a method for the detection and determination of
CC   the metastatic potential of a cell. The method comprises detecting a
CC   gland specific Ets transcription factor (GSEF) gene product in a test
```

CC sample. Detection of a GSEF gene product in the test sample in amount
CC lower than that in a normal cell, is indicative of a cell with high
CC metastatic potential. The method is useful for determining the metastatic
CC potential of a cell, for the diagnosis and prognosis of cancer as well as
CC grading and staging of cancers by detecting GSEF expression in a
CC biological test sample. The method may also be used to monitor patients
CC having a predisposition to develop a particular cancer. GSEF polypeptides
CC are useful for producing antibodies, in cancer diagnosis, prognosis,
CC grading, staging and management of breast and prostate tumours, and in
CC detecting polymorphisms in the sequence. GSEF genes and proteins are also
CC useful in gene therapy. GSEF gene product expression levels can be used
CC in conjunction with any tissue in which an alteration in GSEF gene
CC product expression levels is associated with development of a
CC cancer associated phenotype. Cancers, which can be monitored include
CC cancers of the prostate, cervix, lung and colon, melanomas, colorectal
CC adenocarcinomas, Wilm's tumour, retinoblastoma, sarcomas, myosarcomas,
CC lung carcinomas, leukaemia, and lymphomas. The GSEF gene is located on
CC human chromosome 6, specifically at 6p21.1-6p21.3. The present sequence
CC represents the DNA encoding GSEF.
XX
SQ Sequence 1894 BP; 368 A; 653 C; 571 G; 302 T; 0 other;

alignment_scores:
Quality: 9.09 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_146_154 x AAC83261 ..

Align seq 1/1 to: AAC83261 from: 1 to: 1894

1 AspProMetAspTrpSerProSerAsn 9
||||||||||||||||||||||||||||
851 GATCCCATGGACTGGAGCCCCAGCAAT 877

seq_name: /SIDSL/seqdata/geneseq2/geneseq_pc_NA.seq AC AAC83266

seq_documentation_block:
ID AAC83266 standard; DNA; 3317 BP.
XX
AC AAC83266;
XX
DT 16-MAR-2001 (first entry)
XX
DE Gland-specific Ets transcription factor (GSEF) cDNA sequence.
XX
KW Transcription factor; gland-specific Ets transcription factor; GSEF;
KW metastatic potential; cancer; tumour; metastasis; breast; prostate;
KW leukaemia; lymphoma; sarcoma; melanoma; chromosome 6p21.1-6p21.3; ss.
XX
OS Homo sapiens.
XX
PN WO200070092-A1.
XX
PD 23-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US13173.
XX
PR 14-MAY-1999; 99US-0134112.
XX
PA (CHIR) CHIRON CORP.
XX
PI Kaufmann J, Xin H, Harrowe G;
XX
DR WPI; 2001-041019/05.
DR P-PSDB; AAB49633.
XX
PT Detecting metastatic and potential metastatic cancerous cells, useful
PT for diagnosing, prognosing, grading and staging of cancers by detecting
PT gland-specific Ets transcription factor gene product in a biological
PT sample from a cell -
XX
PS Disclosure; Fig 2; 95pp; English.
XX
CC This invention relates to a method for the detection and determination of
CC the metastatic potential of a cell. The method comprises detecting a
CC gland-specific Ets transcription factor (GSEF) gene product in a test
CC sample. Detection of a GSEF gene product in the test sample in amount
CC lower than that in a normal cell, is indicative of a cell with high
CC metastatic potential. The method is useful for determining the metastatic
CC potential of a cell, for the diagnosis and prognosis of cancer as well as
CC grading and staging of cancers by detecting GSEF expression in a
CC biological test sample. The method may also be used to monitor patients
CC having a predisposition to develop a particular cancer. GSEF polypeptides
CC are useful for producing antibodies, in cancer diagnosis, prognosis,
CC grading, staging and management of breast and prostate tumours, and in
CC detecting polymorphisms in the sequence. GSEF genes and proteins are also
CC useful in gene therapy. GSEF gene product expression levels can be used
CC in conjunction with any tissue in which an alteration in GSEF gene
CC product expression levels is associated with development of a
CC cancer-associated phenotype. Cancers, which can be monitored include
CC cancers of the prostate, cervix, lung and colon, melanomas, colorectal
CC adenocarcinomas, Wilm's tumour, retinoblastoma, sarcomas, myosarcomas,
CC lung carcinomas, leukaemia, and lymphomas. The GSEF gene is located on
CC human chromosome 6, specifically at 6p21.1-6p21.3. The present sequence
CC represents the cDNA encoding GSEF.
XX
SQ Sequence 3317 BP; 710 A; 1026 C; 920 G; 611 T; 0 other;

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_146_154 x AAC83266 ..

Align seq 1/1 to: AAC83266 from: 1 to: 3317

1 AspProMetAspTrpSerProSerAsn 9
|||||||||||||||||||||||||||
2250 GATCCCATGGACTGGAGCCCCAGCAAT 2276

seq name: /SIDSL/seqdata/geneseq2/geneseq_NA1997.DAT:AA167546

seq_documentation_block:
ID AA167546 standard; DNA; 267 BP.
XX
AC AA167546;
XX
DT 11-JUL-1997 (first entry)
XX
DE H. pylori inner membrane protein ORF 26261040.aa.
XX
KW Cytoplasmic; vaccine; prevention, treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator, bacteria; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope; ds.
XX
OS Helicobacter pylori.
XX
FH Key Location/Qualifiers
FT CDS 1..267
FT /*tag= a
FT /note= "no stop codon given"
XX
PN WO9640893-A1.
XX
PD 19-DEC-1996.
XX
PF 06-JUN-1996; 96WO-US09122.
XX
PR 07-JUN-1995; 95US-0487032.

```
XX
PA   (ASTR ) ASTRA AB.
XX
PT   Berglindh OT,  Smith D,  Mellquard BL;
XX
DR   WPI; 1997-052306/05.
DR   P-PSDB; AAW20454.
XX
PT   Helicobacter pylori nucleic acid sequences and related
PT   polypeptide(s)    useful for vaccines to treat or prevent H. pylori
PT   infection, and to detect Helicobacter
XX
PS   Claim 1; Page 230; 1481pp; English.
XX
CC   This sequence encodes a H. pylori inner membrane protein.
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds,
CC   useful as potential H. pylori life cycle activators or inhibitors.
CC   The genomic sequence of H. pylori (ATCC 55679) was determined from
CC   overlapping contigs generated by mechanically shearing the bacterial
CC   DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC   and the predicted coding regions defined by computer evaluation. To
CC   identify likely H. pylori antigens for vaccine development, the amino
CC   acid sequences predicted from various ORF were analysed for significant
CC   homology to other known or exported membrane proteins. Having identified
CC   and determined the sequences of interest, particular regions can be
CC   isolated from H. pylori by PCR amplification for recombinant polypeptide
CC   production, e.g. in E. coli hosts.
XX
SQ   Sequence 267 BP; 88 A; 57 C; 59 G; 63 T; 0 other;


alignment_scores:
              Quality:    6.00              Length:       6
                Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 DS 09-126-945B 2_COPY_146_154 x AAI67546/rev    ..

 Align seq 1/1   to reverse of: AAI67546  from: 1  to: 267

        4 AspTrpSerProSerAsn 9
          ||||||||||||||||||
      266 GATTGGTCTCCTTCTAAT 249

seq_name: /tmp1/peptalign/genesequence_pct_NAI67546.tAA:AAQ57761

seq_documentation_block:
ID   AAQ57761 standard; DNA; 268 BP.
XX
AC   AAQ57761;
XX
DT   15-AUG-1994  (first entry)
XX
DE   Mycobacterium bovis 65kD antigen coding sequence (461..728).
XX
KW   Mycobacterium bovis; species specific detection; probe; primer;
KW   65kD antigen; ss.
XX
OS   Mycobacterium bovis.
XX
PN   EP584024-A.
XX
PD   23-FEB-1994.
XX
PF   12-AUG-1993;   93EP-0420339.
XX
PR   12-AUG-1992;   92FR-0010094.
XX
PA   (MABI/) MABILAT C.
PA   (INMR ) BIO MERIEUX.
XX
PI   Mabilat C,  Pechere J;
XX
DR   WPI; 1994-058092/08.
XX
PT   Mycobacterium DNA fragments - corresp. to 65 kD antigen
PT   sequences, and primers and probes for detecting Mycobacterium spp.
XX
PS   Claim 2; Page 14; 40pp; French.
XX
CC   The region corresponding to nucleotides 438-751 encoding the
CC   M.tuberculosis 65kD antigen is highly conserved among mycobacteria.
CC   DNA fragments having at least 70 per cent homology to this region
CC   and to the corresponding region from other species (e.g. M.bovis
CC   see AAQ57761) are claimed.
XX
SQ   Sequence 268 BP; 57 A; 84 C; 94 G; 33 T; 0 other;


alignment_scores:
              Quality:    6.00              Length:       6
                Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 DS 09-126-945B 2_COPY_146_154 x AAQ57761/rev    ..

 Align seq 1/1   to reverse of: AAQ57761  from: 1  to: 268

        2 ProMetAspTrpSerPro 7
          |||| ||||||||||||
      261 CCGATGGACTGGTCACCC 244

Seq_name: /tmp1/peptalign/genesequence_pct_NAI67546.tAA:AAQ57762

seq_documentation_block:
ID   AAQ57762 standard; DNA; 268 BP.
XX
AC   AAQ57762;
XX
DT   15-AUG-1994  (first entry)
XX
DE   Mycobacterium microti 65kD antigen coding sequence (461..728).
XX
KW   Mycobacterium microti; species specific detection; probe; primer;
KW   65kD antigen; ss.
XX
OS   Mycobacterium microti.
XX
PN   EP584024-A.
XX
PD   23-FEB-1994.
XX
PF   12-AUG-1993;   93EP-0420339.
XX
PR   12-AUG-1992;   92FR-0010094.
XX
PA   (MABI/) MABILAT C.
PA   (INMR ) BIO MERIEUX.
XX
PI   Mabilat C,  Pechere J;
XX
DR   WPI; 1994-058092/08.
XX
PT   Mycobacterium DNA fragments - corresp. to 65 kD antigen
PT   sequences, and primers and probes for detecting Mycobacterium spp.
XX
PS   Claim 2; Page 15; 40pp; French.
XX
CC   The region corresponding to nucleotides 438-751 encoding the
CC   M.tuberculosis 65kD antigen is highly conserved among mycobacteria.
CC   DNA fragments having at least 70 per cent homology to this region
CC   and to the corresponding region from other species (e.g. M.microti
CC   see AAQ57762) are claimed.
```

```
XX
SQ   Sequence 268 BP; 57 A; 84 C; 94 G; 33 T; 0 other;


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_146_154 x AAQ57762/rev

 Align seq 1/1  to reverse of: AAQ57762  from: 1  to: 268

       2 ProMetAspTrpSerPro 7
         |||||||||||||||||||
     261 CCGATGGACTGGTCACCC 244

seq_name: [SIDS] [Publication Specific Data Set] [AAQ57762]

seq_documentation_block:
ID   AAQ57763 standard; DNA; 268 BP.
XX
AC   AAQ57763;
XX
DT   15-AUG-1994  (first entry)
XX
DE   Mycobacterium africanum 65kD antigen coding sequence (451-728).
XX
KW   Mycobacterium africanum; species specific detection; probe; primer;
KW   65kD antigen; ss.
XX
OS   Mycobacterium africanum.
XX
PN   EP584023-A.
XX
PD   23-FEB-1994.
XX
PF   12-AUG-1993;   93EP-0420339.
XX
PR   12-AUG-1992;   92FR-0010094.
XX
PA   (MABI/) MABILAT C.
PA   (INMR ) BIO MERIEUX.
XX
PI   Mabilat C;
XX
DR   WPI; 1994-098892/08.
XX
PT   Mycobacterium DNA fragments - corresp. to 65 kD antigen
PT   sequences, and primers and probes for detecting Mycobacterium spp.
XX
PS   Claim 2; Page 16; 40pp; French.
XX
CC   The region corresponding to nucleotides 438-751 encoding the
CC   M.tuberculosis 65kD antigen is highly conserved among mycobacteria.
CC   DNA fragments having at least 70 per cent homology to this region
CC   and to the corresponding region from other species (e.g. M.africanum,
CC   see AAQ57763) are claimed.
XX
SQ   Sequence 268 BP; 58 A; 84 C; 93 G; 33 T; 0 other;


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_146_154 x AAQ57763/rev   ..

 Align seq 1/1  to reverse of: AAQ57763  from: 1  to: 268

       2 ProMetAspTrpSerPro 7
         |||||||||||||||||||
     261 CCGATGGACTGGTCACCC 244

seq_name: [SIDS] [Publication Specific Data Set] [AAQ57763]

seq_documentation_block:
ID   AAQ06505 standard; DNA; 442 BP.
XX
AC   AAQ06505;
XX
DT   22-FEB-1991  (first entry)
XX
DE   Encodes protein XXI homologous to mycobacterial antigen.
XX
KW   mycobacterial antigen; tuberculosis; actinomycetales; ds.
XX
OS   Mycobacterium tortuitum.
XX
PN   WO9012879-A.
XX
PD   01-NOV-1990.
XX
PF   13-APR-1990;   90WO-FR00274.
XX
PR   17-APR-1989;   89FR-000505?.
XX
PA   (INRM ) INSERM INST NAT SANTE.
PA   (INSP ) INST PASTEUR.
XX
PI   Rance A,  Grandchamp Bestaux B,  Levy Frebault V,  Grosset B;
XX
DR   WPI; 1990-348478/46.
DR   P-PSDB; AAR07688.
XX
PT   Nucleotide sequences of actinomycetales - used as primers for
PT   synthesis of DNA of actinomycetales
XX
PS   Claim 5; Page 32; 61pp; French.
XX
CC   This sequence is similar to the mycobacterial gene which encodes a
CC   65KD antigen. Regions within this sequence can be used as primers to
CC   amplify mycobacterial genes and detect mycobacteria. Antibodies can
CC   be induced by immunisation with the translation product from this
CC   sequence or its fragments. These can also be used to detect
CC   mycobacteria in assays such as ELISA or RIA procedures.
CC   See also AAQ06506-Q06523 and AAK08336.
XX
SQ   Sequence 442 BP; 75 A; 108 C; 116 G; 43 T; 0 other;


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_146_154 x AAQ06505/rev   ..

 Align seq 1/1  to reverse of: AAQ06505  from: 1  to: 442

       2 ProMetAspTrpSerPro 7
         |||||||||||||||||||
     278 CCGATGGACTGGTCACCC 261

seq_name: [SIDS] [Publication Specific Data Set] [AAQ06505]

seq_documentation_block:
ID   AAQ06507 standard; DNA; 442 BP.
XX
AC   AAQ06507;
XX
DT   22-FEB-1991  (first entry)
```

```
XX
DE   Encodes protein IV homologous to mycobacterial antigen.
XX
KW   mycobacterial antigen; actinomycetales; ds.
XX
OS   Mycobacterium tuberculosis.
XX
PN   WO9012875 A.
XX
PD   01 NOV 1990.
XX
PF   13 APR 1990;   90WO FR00274.
XX
PR   17 APR 1989;   89FR 0005057.
XX
PA   (INRM ) INSERM INST NAT SANTE.
PA   (INSP ) INST PASTEUR.
XX
PI   Rance A,  Grandchamp-Desraux B,  Levy Prelaudt V,  Clerget B,
XX
DR   WPI; 1990-348478/46.
XX
PT   Nucleotide sequences of actinomycetales - used as primers for
PT   synthesis of DNA of actinomycetales.
XX
PS   Claim 9; Page 84; 61pp; French.
XX
CC   This sequence corresponds to the mycobacterial gene which encodes a
CC   65KD antigen of the group of tuberculosis bacteria. Regions within
CC   this sequence can be used as primers to amplify mycobacterial genes
CC   and detect mycobacteria. Antibodies can be induced by immunisation
CC   with the DNA sequence or its fragments. These can also be used to
CC   detect mycobacteria in assays such as ELISA or RIA procedures.
CC   See also AAQ06505-Q06506, AAQ06508-Q06523 and AAR08436.
XX
SQ   Sequence 342 BP; 75 A; 104 C; 122 G; 42 T; 0 other;


alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2_COPY_146_154 x AAQ06507/rev   ..

 Align seq 1/1  to reverse of: AAQ06507  from: 1  to: 342

        2 ProMetAspTrpSerPro 7
          ||||||||||||||||||
      278 CCGATGGACTGGTCACCC 261

seq_name: "/cds1/pending/geneseq/geneseqp-NA1997.ofA:AAI67868"

seq_documentation_block:
ID   AAI67868 standard; DNA; 432 BP.
XX
AC   AAI67868;
XX
DT   14-JUL-1997  (first entry)
XX
DE   H. pylori cytoplasmic protein ORF 01xe21717orf19.
XX
KW   Vaccine; prevention; treatment; infection; identification;
KW   binding compound; bacterium; life cycle; activator; bacteria;
KW   inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
KW   cytoplasmic; ds.
XX
OS   Helicobacter pylori.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..432
FT                   /*tag= a
FT                   /note= "no stop codon given"
XX
PN   WO9640908-A1.
XX
PD   19-DEC-1996.
XX
PF   06-JUN-1996;   96WO-US09122.
XX
PR   01-APR-1996;   96US-0640405.
PR   07-JUN-1995;   95US-0487032.
XX
PA   (ASTR ) ASTRA AB.
XX
PI   Berglindh T-I,  Smith D,  Melhaerd B-L;
XX
DR   WPI; 1997-05246/05.
DR   P-PSDB; AAW20615.
XX
PT   Helicobacter pylori nucleic acid sequences and related
PT   polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT   infection, and to detect Helicobacter
XX
PS   Claim 9; Page 757; 1481pp; English.
XX
CC   The present sequence encodes a Helicobacter pylori cytoplasmic
CC   protein.
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds,
CC   useful as potential H. pylori life cycle activators or inhibitors.
CC   The genomic sequence of H. pylori (ATCC 55679) was determined from
CC   overlapping contigs generated by mechanically shearing the
CC   bacterial DNA. The sequences were analysed for ORF of at least 180
CC   nucleotides, and the predicted coding regions defined by computer
CC   evaluation. To identify likely H. pylori antigens for vaccine
CC   development, the amino acid sequences predicted from various ORF
CC   were analysed for significant homology to other known or exported
CC   membrane proteins. Having identified and determined the sequences
CC   of interest, particular regions can be isolated from H. pylori by
CC   PCR amplification for recombinant polypeptide production, esp. in
CC   E. coli hosts.
XX
SQ   Sequence 432 BP; 109 A; 87 C; 86 G; 120 T; 0 other;


alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2_COPY_146_154 x AAI67868/rev   ..

 Align seq 1/1  to reverse of: AAI67868  from: 1  to: 432

        4 AspTrpSerProSerAsn 9
          |||||||||||||||||||
      18 GATTGGTCACCTTCTAAT 1

seq_name: "/cds1/pending/geneseq/geneseqp-NA1997.ofA:AAI67864"

seq_documentation_block:
ID   AAI67864 standard; DNA; 681 BP.
XX
AC   AAI67864;
XX
DT   14-JUL-1997  (first entry)
XX
DE   H. pylori inner membrane protein ORF 01xe21717orf18.
XX
KW   Vaccine; prevention; treatment; infection; identification;
KW   binding compound; bacterium; life cycle; activator; bacteria;
KW   inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
KW   inner membrane; ds.
```

```
XX
OS   Helicobacter pylori.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..681
FT                   /*tag   a
FT                   /note= "no stop codon given"
XX
PN   WO9640893-A1.
XX
PD   19-DEC-1996.
XX
PF   06-JUN-1996;    96WO-US09122.
XX
PR   01-APR-1996;    96US-0630405.
PR   07-JUN-1995;    95US-0487042.
XX
PA   (ASIR ) ASTRA AB.
XX
PI   Berglindh CF,  Smith D,  Mellgaerd BL;
XX
DR   WPI; 1997-052306/05.
DR   P-PSDB; AAW20611.
XX
PT   Helicobacter pylori nucleic acid sequences and related
PT   polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT   infection, and to detect Helicobacter
XX
PS   Disclosure; Page 755; 1481pp; English.
XX
CC   The present sequence encodes a Helicobacter pylori inner membrane
CC   protein.
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds,
CC   useful as potential H. pylori life cycle activators or inhibitors.
CC   The genomic sequence of H. pylori (ATCC 55679) was determined from
CC   overlapping contigs generated by mechanically shearing the
CC   bacterial DNA. The sequences were analysed for ORF of at least 180
CC   nucleotides, and the predicted coding regions defined by computer
CC   evaluation. To identify likely H. pylori antigens for vaccine
CC   development, the amino acid sequences predicted from various ORF
CC   were analysed for significant homology to other known or reported
CC   membrane proteins. Having identified and determined the sequences
CC   of interest, particular regions can be isolated from H. pylori by
CC   PCR amplification for recombinant polypeptide production, e.g. in
CC   E. coli hosts.
XX
SQ   Sequence 681 BP; 214 A; 128 C; 166 G; 173 T; 0 other;


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_146_154 x AA167864/rev   ..

 Align seq 1/1  to reverse of: AAT67864  from: 1  to: 681

       4 AspTrpSerProSerAsn 9
         |||||||||||||||||
     680 GATTGGTCTCCTTCTAAT 663

seq_name: /SIDSI/seqdata/geneseq/geneseqn/NA2001.DAT:AAF25022

seq_documentation_block:
ID   AAF25022 standard; DNA; 888 BP.
XX
AC   AAF25022;
XX
DT   30-APR-2001  (first entry)
XX
DE   Nucleotide sequence of Hsp65-E7 fusion protein.
XX
KW   Heat shock protein; Hsp; Th1 response; Th1 cell; CD4+ T lymphocyte cell;
KW   lymphocyte; Hsp65, Hsp40, Hsp10, Hsp60, Hsp71, microbial pathogen;
KW   E7 protein; ss.
XX
OS   Synthetic.
OS   Mycobacterium bovis.
OS   Human papillomavirus
XX
FH   Key             Location/Qualifiers
FT   CDS             1..888
FT                   /*tag   a
FT                   /product  "Hsp65-E7 fusion protein"
XX
PN   WO200104344-A2.
XX
PD   18-JAN-2001.
XX
PF   10-JUL-2000; 2000WO-US18828.
XX
PR   08-JUL-1999;   99US-0143257.
XX
PA   (STRE-) STRESSGEN BIOTECHNOLOGIES CORP.
XX
PI   Siegel M,  Chi NR,  Mizzen LA;
XX
DR   WPI; 2001-148461/14.
DR   P-PSDB; AAB31615.
XX
PT   Screening for compounds that stimulate Th1-like responses in CD4+ T
PT   lymphocyte cells -
XX
PS   Example 12; Fig 11; 88pp; English.
XX
CC   The present sequence encodes a fusion protein comprising a Mycobacterium
CC   bovis heat shock protein (Hsp) 65 fused at its C terminal to a HPV16 E7
CC   protein. The fusion protein is used in the method of the invention. The
CC   specification describes a method of determining whether a compound
CC   stimulates a Th1-like response. Th1 cells are a subset of CD4+
CC   T lymphocyte cells. The method comprises contacting naive lymphocytes
CC   in vitro with a fusion protein comprising at least a fragment of Hsp,
CC   and then detecting the Th1-like response exhibited by the cell sample.
CC   The proteins which may be used in the method of the invention are Hsp65,
CC   Hsp40, Hsp10, Hsp60, and Hsp71. The method may be used to identify
CC   compounds that stimulate Th1-like responses in response to microbial
CC   pathogens.
XX
SQ   Sequence 888 BP; 224 A; 235 C; 272 G; 157 T; 0 other;


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_146_154 x AAF25022/rev   ..

 Align seq 1/1  to reverse of: AAF25022  from: 1  to: 888

       2 ProMetAspIrpSerPro 7
         |||||||||||||||||
     470 CCGATGGACTGGTCACCC 453

seq_name: /SIDSI/seqdata/geneseq/geneseqn/NA2001.DAT:AAC61507

seq_documentation_block:
ID   AAC61507 standard; DNA; 915 BP.
XX
AC   AAC61507;
XX
DT   19-FEB-2001  (first entry)
```

```
XX
DE   Sequence of VP1 region of enterovirus serotype E34 strain DN 19.
XX
KW   VP1 gene; enterovirus; serotype typing; virus detection; ss.
XX
OS   Human enterovirus.
XX
PN   WO200058524 A2.
XX
PD   05 OCT 2000.
XX
PF   24 MAR 2000; 2000WO US07828.
XX
PR   31 MAR 1999;   99US 0127464.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Oberste S, Maher K, Kilpatrick DK, Pallansch MA.
XX
DR   WPI; 2000 664930/64.
XX
PT   Detecting the presence of an enterovirus and typing an enterovirus in a
PT   sample for diagnostic use, involves amplifying viral RNA and sequencing
PT   the resulting amplicons
XX
PS   Example 1; Page 94; 97pp; English.
XX
CC   AAC61454-C61509 represent VP1 region sequences from different human
CC   non polio enterovirus serotypes and strains. The sequences can be
CC   detected using the method of the invention. The specification
CC   describes a method for detecting the presence of an enterovirus in
CC   a sample. The method comprises purifying RNA from the sample, reverse
CC   transcribing the RNA, amplifying a portion of the VP1 gene of an
CC   enterovirus from the cDNA, and then detecting the amplicon, which
CC   indicates presence of the enterovirus in the sample. The method is
CC   useful for detecting the presence of an enterovirus and for typing
CC   the serotype of the enterovirus in a clinical sample.
XX
SQ   Sequence 915 BP; 281 A; 227 C; 192 G; 215 T; 0 other;


alignment scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2 COPY 146 154 x AAC61507   ..

 Align seq 1/1  to: AAC61507  from: 1  to: 915

        4 MetAspTrpSerProSer 8
          ||||||||||||||||||
      692 ATGGATTGGTCACCATCA 709

Seq name:  /cgi-bin/patbin/geneseq_nucseqs_NA.pl?...DAT:AAZ20163

seq_documentation block:
ID   AAZ20163 standard; cDNA; 1459 BP.
XX
AC   AAZ20163;
XX
DT   05-JAN-2000  (first entry)
XX
DE   Bovine pregnancy associated glycoprotein boPAG4 cDNA.
XX
KW   PAG; boPAG4; pregnancy associated glycoprotein; cattle; bovine;
KW   early pregnancy diagnosis; ss.
XX
OS   Bos taurus.
XX
PN   WO9947944 A2.
XX
PD   23-SEP-1999
XX
PF   19-MAR-1999;   99WO-US06038.
XX
PR   20-MAR-1998;   98US-0078784.
PR   28-OCT-1998;   98US-0106188.
XX
PA   (BMGR ) BNTV MISSOURI.
XX
PI   Roberts RM, Green JA, Xie S;
XX
DR   WPI; 1999-601142/51.
XX
PT   New bovine polypeptides useful for early diagnosis of pregnancy
XX
PS   Claim 69; Page 92; 136pp; English.
XX
CC   This is the nucleotide sequence of cDNA forming part of the coding
CC   region encoding bovine pregnancy associated glycoprotein (PAG)
CC   boPAG4 (see AAY32036). PAGs are structurally related to pepsins,
CC   thought to be restricted to ungulates, and are specifically
CC   expressed in the chorion/trophectoderm of the placenta. PAGs (see
CC   AAY32035-38) are highly diverse in sequence, with regions of
CC   hypervariability largely confined to surface exposed loops.
CC   Selected PAGs (e.g. boPAG4) that are products of invasive binucleate
CC   cells, expressed highly in early pregnancy at the time of
CC   trophoblast invasion, and expressed weakly, if at all, in late
CC   gestation, are useful in the early diagnosis of pregnancy.
CC   Immunoassays for detecting such PAGs are disclosed. Identification
CC   of PAGs allows detection of pregnancy in cattle, goat, sheep,
CC   rhinoceros, horse, cat, dog and human (all claimed). Isolated PAG
CC   nucleic acids can be used in the recombinant production of the PAGs,
CC   or to generate probes and primers.
XX
SQ   Sequence 1459 BP; 447 A; 430 C; 423 G; 359 T; 0 other;


alignment scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_146_154 x AAZ20163/rev   ..

 Align seq 1/1  to reverse of: AAZ20163  from: 1  to: 1459

        2 ProMetAspTrpSerPro 7
          |||||||||||||||||
    1269 CCCATGGACTGGAGCCCA 1252
```

```
OM of: US-09-126-945B-2_COPY_146_154 *** [sec] Title: NA *

Date: Sep 14, 2001  9:56 PM

About: Results were provided by the Center for Biological
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL frame+ p2n_mm_bd -DEV xlp
-Q /usr2/prodata/2/ina/5A_COMB/seq_gb_10/gb_10_5A_1.seq,/usr2/pro_apps_port/data1_1/seq
-DB Issued_Patents_NA -QFMT fastap -SUFFIX olip2n.rni
-GAPOP 4.500 -GAPEXT 0.050 -MINMATCH 0.100 -L 0.000 0.0000
-LOOPEXT 0.000 -QGAPOP 4.500 -QGAPEXT 0.050 -XGAPOP 60.000
-XGAPEXT 60.000 -PGAPOP 6.000 -PGAPEXT 7.000 -YGAPOP 60.000
-YGAPEXT 60.000 -DELOP 6.000 -DELEXT 7.000 -START 1
-MATRIX oligo -TRANS human40.cdi -LIST 45 -ALIGN 200
-THR_SCORE quality -THR_MIN 1 -ALIGN 15 -MODE local -OUTFMT ;;s
-NORM ext -MINLEN 0 -MAXLEN 2000000000
-USER US09126945B_COPY_146_154 -NCPU 6 -TPU 1 -O NULL -N1_XLIXY
-WAIT -THREADS 1

Search information block:
Query: US-09-126-945B-2_COPY_146_154
Query length: 9
Database: Issued_Patents_NA:*
Database sequences: 324599
Database length: 94655562
Search time (sec): 322.720000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000


score_list:
Sequence            Strd Orig    ZScore   EScore Len  ! Documentation
/opn2_6/prodata/2/ina/5A_COMB seq-US-08-105-168B-1      6.00   125.47   22.58      268  !
/opn2_6/prodata/2/ina/5A_COMB seq-US-08-105-168B-2      6.00   125.47   22.58      268  !
/opn2_6/prodata/2/ina/5A_COMB seq-US-08-125-168B-1      6.00   125.47   22.58      268  !
/opn2_6/prodata/2/ina/5B_COMB seq-US-08-698-448-1       6.00   125.47   22.58      268  !
/opn2_6/prodata/2/ina/5A_COMB seq-US-08-664-414-1       6.00   125.47   22.58      268  !
/opn2_6/prodata/2/ina/5B_COMB seq-US-08-698-448-2       6.00   125.47   22.58      268  !
/opn2_6/prodata/2/ina/5A_COMB seq-US-08-165-168B-2      6.00   124.11   22.96      211  !
/opn2_6/prodata/2/ina/5B_COMB seq-US-08-698-448-24      6.00   124.11   22.96      211  !
/opn2_6/prodata/2/ina/5B_COMB seq-US-08-474-620A-2      6.00   124.84   22.82      414  !
/opn2_6/prodata/2/ina/5B_COMB seq-US-08-474-620A-22     6.00   124.84   22.82      414  !
/opn2_6/prodata/2/ina/5B_COMB seq-US-08-474-620A-2      6.00   124.84   22.82      414  !
/opn2_6/prodata/2/ina/5B_COMB seq-US-08-474-620A-1      6.00   124.84   22.82      414  !
/opn2_6/prodata/2/ina/5A_COMB seq-US-08-105-168B-22     6.00   123.69   23.07      450  !
/opn2_6/prodata/2/ina/5B_COMB seq-US-08-698-448-22      6.00   123.69   23.07      450  !
/opn2_6/prodata/2/ina/5B_COMB seq-US-08-636-450-1       6.00   120.66   23.63      455  !
/opn2_6/prodata/2/ina/5B_COMB seq-US-08-665-604A-14     5.00   111.68   90.04       70  !
/opn2_6/prodata/2/ina/5B_COMB seq-US-08-665-604A-13     5.00   111.68   90.04       70  !
/opn2_6/prodata/2/ina/5B_COMB seq-US-08-672-125-20      5.00   111.68   90.04       70  !
/usr2/prodata/2/ina/5B_COMB seq-PCT-US94-004311-20      5.00   111.68   90.04       70  !
/usr2/prodata/2/ina/5B_COMB seq-US-07-916-604A-22       5.00   101.82   113.20      221  !
/usr2/prodata/2/ina/5B_COMB seq-US-08-421-991B-1        5.00   101.15   115.62      500  !
/usr2/prodata/2/ina/5B_COMB seq-US-09-004-415-20        5.00    99.54   116.11      498  !
/usr2/prodata/2/ina/5B_COMB seq-US-08-008-116-20        5.00    99.04   116.26      728  !
/usr2/prodata/2/ina/5B_COMB seq-US-08-008-116-20        5.00    99.04   116.26      728  !
/usr2/prodata/2/ina/5B_COMB seq-US-08-004-276-20-1      5.00    97.71   116.72      992  !
/usr2/prodata/2/ina/5A_COMB seq-US-08-105-168B-22       5.00    97.14   116.91      921  !
/usr2/prodata/2/ina/5A_COMB seq-US-08-764-620A-20       5.00    97.14   116.91      921  !
/usr2/prodata/2/ina/5B_COMB seq-PCT-US95-010044-20      5.00    97.14   116.91      921  !
/usr2/prodata/2/ina/5B_COMB seq-US-08-764-620A-1        5.00    96.99   116.96      921  !
/usr2/prodata/2/ina/5B_COMB seq-US-08-560-736-1         5.00    96.99   116.96     1000  !
/usr2/prodata/2/ina/5A_COMB seq-US-08-943-620A-1        5.00    96.41   117.15     1047  !
/usr2/prodata/2/ina/5B_COMB seq-US-08-054-011-1         5.00    96.41   117.15     1047  !
/usr2/prodata/2/ina/5B_COMB seq-US-09-159-417-1         5.00    96.41   117.15     1047  !
/usr2/prodata/2/ina/5B_COMB seq-US-09-012-971-6         5.00    96.41   117.15     1047  !
/usr2/prodata/2/ina/6B_COMB seq-US-08-768-473-3         5.00    95.87   117.33     1174  !
/usr2/prodata/2/ina/6A_COMB seq-US-09-082-270-1         5.00    95.76   117.37     1195  !
/usr2/prodata/2/ina/5A_COMB seq-US-08-759-621-16        5.00    95.23   117.54     1292  !
/usr2/prodata/2/ina/5A_COMB seq-US-08-596-171A-17       5.00    95.18   117.56     1245  !
/usr2/prodata/2/ina/5B_COMB.seq-US-08-646-590B-17       5.00    95.18   117.56     1245  !
/usr2/prodata/2/ina/5A_COMB.seq-US-09-060-208-17        5.00    95.18   117.56     1245  !
/usr2/prodata/2/ina/5B_COMB.seq-US-09-112-184-17        5.00    95.18   117.56     1245  !
/usr2/prodata/2/ina/5B_COMB.seq-US-08-671-975A-1        5.00    95.18   117.56     1306  !
/usr2/prodata/2/ina/6B_COMB.seq-US-09-143-954-3         5.00    93.50   118.16     1674  !
/usr2/prodata/2/ina/5B_COMB.seq-US-08-365-486A-11       5.00    93.45   118.18     1679  !
/usr2/prodata/2/ina/6B_COMB.seq-US-08-880-342-11        5.00    93.45   118.18     1679  !

>>seq_num: /usr2/prodata/2/ina/5A_COMB.seq:US-08-105-168B-1

Seq_documentation_block:
; Sequence 1, Application US/08105168B
; Patent No. 5589585
;  GENERAL INFORMATION:
;    APPLICANT:  MABILAT et al.
;    TITLE OF INVENTION:  DNA FRAGMENTS OF MYCOBACTERIA, AMPLIFICATION
;    TITLE OF INVENTION:  PRIMERS, HYBRIDIZATION PROBES, REAGENTS AND METHODS FOR THE
;    TITLE OF INVENTION:  MYCOBACTERIA
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Olitt & Berridge
;      STREET:  700 South Washington Street, Suite 400
;      CITY:  Alexandria,
;      STATE:  Virginia
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" DS/HD
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS DOS 6.1
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      FILING DATE:  August 12, 1993
;      APPLICATION NUMBER:  US/08/105,168B
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR9210094
;      FILING DATE:  August 8, 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  William P. Berridge
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 28835
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6400
;      TELEFAX:  (703) 836-2787
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  268 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:
;    ANTI-SENSE:
;    ORIGINAL SOURCE:
;      ORGANISM:  Mycobacterium bovis
;      STRAIN:
;      INDIVIDUAL OR ISOLATE:
;    POSITION IN GENOME:
;      MAP POSITION:
;    FEATURE:
;      NAME/KEY:
;      LOCATION:  471-738
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:
>>> No HSPs [Abs]


Alignment_scores:
            Quality:    6.00              Length:      6
              Ratio:   1.000               Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000
```

```
alignment block:
 US-09-126-945B-2 copy 146-154 x US-08-105-168B-1/rev

 Align seq 1/1  to reverse of: US-08-105-168B-1  from: 1  to: 268

        2 ProMetAspTrpSerPro 7
          |||||||||||||||||||
      261 CCGATGGACTGGTCACCC 244

seq name: /tmp/_fpt-data/_tca_5A_CoMB.seq:US-08-105-168B-2

seq documentation block:
; Sequence 2, Application US/08105168B
; Patent No. 5589585
;  GENERAL INFORMATION:
;     APPLICANT:  MABILAT et al.
;     TITLE OF INVENTION:  DNA FRAGMENTS OF MYCOBACTERIA, AMPLIFICATION
;     TITLE OF INVENTION:  PRIMERS, HYBRIDIZATION PROBES, REAGENTS AND METHOD FOR THE DET
;     TITLE OF INVENTION:  MYCOBACTERIA
;     NUMBER OF SEQUENCES:  23
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Olitt & Berridge
;       STREET:  700 South Washington Street, Suite 400
;       CITY:  Alexandria,
;       STATE:  Virginia
;       ZIP:  22314
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5" DS/HD
;       COMPUTER:  IBM compatible
;       OPERATING SYSTEM:  MS DOS 3.1
;       SOFTWARE:  Wordperfect
;     CURRENT APPLICATION DATA:
;       FILING DATE:  August 12, 1993
;       APPLICATION NUMBER:  US/08/105,168B
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  FR9210094
;       FILING DATE:  August 8, 1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  William P. Berridge
;       REGISTRATION NUMBER:  30,024
;       REFERENCE/DOCKET NUMBER:  WPB 28835
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 836 6400
;       TELEFAX:  (703) 836 2787
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  268 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single stranded
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:
;     ANTI-SENSE:
;     ORIGINAL SOURCE:
;       ORGANISM:  Mycobacterium microti
;       STRAIN:
;       INDIVIDUAL ISOLATE:
;     POSITION IN GENOME:
;       CHROMOSOME/SEGMENT:
;       MAP POSITION:
;     FEATURE:
;       NAME/KEY:
;       LOCATION:  461..728
;       IDENTIFICATION METHOD:
;       OTHER INFORMATION:
US-08-105-168B-2


alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000
```

```
alignment block:
 US-09-126-945B-2 copy 146-154 x US-08-105-168B-2/rev

 Align seq 1/1  to reverse of: US-08-105-168B-2  from: 1  to: 268

        2 ProMetAspTrpSerPro 7
          |||||||||||||||||||
      261 CCGATGGACTGGTCACCC 244

seq name: /tmp/_fpt-data/_tca_5A_CoMB.seq:US-08-105-168B-3

seq documentation block:
; Sequence 3, Application US/08105168B
; Patent No. 5589585
;  GENERAL INFORMATION:
;     APPLICANT:  MABILAT et al.
;     TITLE OF INVENTION:  DNA FRAGMENTS OF MYCOBACTERIA, AMPLIFICATION
;     TITLE OF INVENTION:  PRIMERS, HYBRIDIZATION PROBES, REAGENTS AND METHOD FOR THE
;     TITLE OF INVENTION:  MYCOBACTERIA
;     NUMBER OF SEQUENCES:  23
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Olitt & Berridge
;       STREET:  700 South Washington Street, Suite 400
;       CITY:  Alexandria,
;       STATE:  Virginia
;       ZIP:  22314
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5" DS/HD
;       COMPUTER:  IBM compatible
;       OPERATING SYSTEM:  MS DOS 3.1
;       SOFTWARE:  Wordperfect
;     CURRENT APPLICATION DATA:
;       FILING DATE:  August 12, 1993
;       APPLICATION NUMBER:  US/08/105,168B
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  FR9210094
;       FILING DATE:  August 8, 1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  William P. Berridge
;       REGISTRATION NUMBER:  30,024
;       REFERENCE/DOCKET NUMBER:  WPB 28835
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 836 6400
;       TELEFAX:  (703) 836-2787
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  268 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single stranded
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:
;     ANTI-SENSE:
;     ORIGINAL SOURCE:
;       ORGANISM:  Mycobacterium africanum
;       STRAIN:
;       INDIVIDUAL ISOLATE:
;     POSITION IN GENOME:
;       CHROMOSOME/SEGMENT:
;       MAP POSITION:
;     FEATURE:
;       NAME/KEY:
;       LOCATION:  461..728
;       IDENTIFICATION METHOD:
;       OTHER INFORMATION:
US-08-105-168B-3


alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
```

```
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US 09-126-945B-2_COPY_146_154 x US-08-105-168A-3  from: 1  to: 268

 Align Seq 1/1  to reverse of: US-08-105-168A-3  from: 1  to: 268

       2 ProMetAspTrpSerPro 7
         |||||||||||||||||||
     261 CCGATGGACTGGTCACCC 244

seq_name: /usr2/prd/data/US_08_MPB_SEQ/US-08-698-948-2

seq_documentation_block:
; Sequence 1, Application US/08698948
; Patent No. 5849901
;  GENERAL INFORMATION:
;     APPLICANT:  MABILAT et al.
;     TITLE OF INVENTION:  DNA FRAGMENTS OF MYCOBACTERIA, AMPLIFICATION
;     TITLE OF INVENTION:  PRIMERS, HYBRIDIZATION PROBES, REAGENTS AND METHOD FOR THE DET
;     TITLE OF INVENTION:  MYCOBACTERIA
;     NUMBER OF SEQUENCES:  23
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Olitt & Berridge
;       STREET:  700 South Washington Street, Suite 400
;       CITY:  Alexandria,
;       STATE:  Virginia
;       ZIP:  22314
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5" DS/HD
;       COMPUTER:  IBM compatible
;       OPERATING SYSTEM:  MS DOS 3.1
;       SOFTWARE:  Wordperfect
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/698,948
;       FILING DATE:  August 16, 1996
;       CLASSIFICATION:  536
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/105,168
;       FILING DATE:  August 12, 1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  FR9210094
;       FILING DATE:  August 8, 1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  William P. Berridge
;       REGISTRATION NUMBER:  30,024
;       REFERENCE/DOCKET NUMBER:  WPB 28835A
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 836-6400
;       TELEFAX:  (703) 836-2787
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  268 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single-stranded
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:
;     ANTI-SENSE:
;     ORIGINAL SOURCE:
;       ORGANISM:  Mycobacterium bovis
;       STRAIN:
;       INDIVIDUAL ISOLATE:
;     POSITION IN GENOME:
;       CHROMOSOME/SEGMENT:
;       MAP POSITION:
;     FEATURE:
;       NAME/KEY:
;       LOCATION:  461-728
;       IDENTIFICATION METHOD:
;       OTHER INFORMATION:
US-08-698-948-1

alignment_scores:
            Quality:    6.00              Length:      6
              Ratio:    1.00                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-08-698-948-1 x US-09-126-945B-2_COPY_146_154  from: 1  to: 268

 Align Seq 1/1  to reverse of: US-08-698-948-1  from: 1  to: 268

       2 ProMetAspTrpSerPro 7
         |||||||||||||||||||
     261 CCGATGGACTGGTCACCC 244

seq_name: /usr2/prd/data/US_08_MPB_SEQ/US-08-698-948-2

seq_documentation_block:
; Sequence 2, Application US/08698948
; Patent No. 5849901
;  GENERAL INFORMATION:
;     APPLICANT:  MABILAT et al.
;     TITLE OF INVENTION:  DNA FRAGMENTS OF MYCOBACTERIA, AMPLIFICATION
;     TITLE OF INVENTION:  PRIMERS, HYBRIDIZATION PROBES, REAGENTS AND METHOD FOR THE
;     TITLE OF INVENTION:  MYCOBACTERIA
;     NUMBER OF SEQUENCES:  23
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Olitt & Berridge
;       STREET:  700 South Washington Street, Suite 400
;       CITY:  Alexandria,
;       STATE:  Virginia
;       ZIP:  22314
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5" DS/HD
;       COMPUTER:  IBM compatible
;       OPERATING SYSTEM:  MS DOS 3.1
;       SOFTWARE:  Wordperfect
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/698,948
;       FILING DATE:  August 16, 1996
;       CLASSIFICATION:  536
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/105,168
;       FILING DATE:  August 12, 1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  FR9210094
;       FILING DATE:  August 8, 1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  William P. Berridge
;       REGISTRATION NUMBER:  30,024
;       REFERENCE/DOCKET NUMBER:  WPB 28835A
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 836-6400
;       TELEFAX:  (703) 836-2787
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  268 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single-stranded
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:
;     ANTI-SENSE:
;     ORIGINAL SOURCE:
;       ORGANISM:  Mycobacterium microti
;       STRAIN:
;       INDIVIDUAL ISOLATE:
;     POSITION IN GENOME:
;       CHROMOSOME/SEGMENT:
;       MAP POSITION:
;     FEATURE:
;       NAME/KEY:
;       LOCATION:  461-728
```

```
;       IDENTIFICATION METHOD:
;       OTHER INFORMATION:
US-08-698-948-2


alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_copy_146_154 x US-08-698-948-2/rev   ..

 Align seq 1/1  to reverse of: US-08-698-948-2  from: 1  to: 268

        2 ProMetAspIrpSerPro 7
          |||||||||||||||||
      261 CCGATGGACTGGTCACCC 244

Seq name: /cgn2_6/ptodata/2/iia/1617-MB_us-09-126-945b-2.seq

seq documentation block:
; Sequence 4, Application US/08698948
; Patent No. 5849901
;  GENERAL INFORMATION:
;     APPLICANT:  MABILAT et al.
;     TITLE OF INVENTION:  DNA FRAGMENTS OF MYCOBACTERIA, AMPLIFICATION
;     TITLE OF INVENTION:  PRIMERS, HYBRIDIZATION PROBES, REAGENTS AND METHOD FOR THE DET
;     TITLE OF INVENTION:  MYCOBACTERIA
;     NUMBER OF SEQUENCES:  23
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Olitt & Berridge
;       STREET:  700 South Washington Street, Suite 300
;       CITY:  Alexandria,
;       STATE:  Virginia
;       ZIP:  22314
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5" DS/HD
;       COMPUTER:  IBM compatible
;       OPERATING SYSTEM:  MS DOS 3.1
;       SOFTWARE:  Wordperfect
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/698,948
;       FILING DATE:  August 16, 1996
;       CLASSIFICATION:  536
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/105,168
;       FILING DATE:  August 12, 1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  FR9210094
;       FILING DATE:  August 8, 1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  William P. Berridge
;       REGISTRATION NUMBER:  30,024
;       REFERENCE/DOCKET NUMBER:  WPB 28835A
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 836 6400
;       TELEFAX:  (703) 836 2787
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  268 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single stranded
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:
;     ANTI SENSE:
;     ORIGINAL SOURCE:
;       ORGANISM:  Mycobacterium africanum
;       STRAIN:
;       INDIVIDUAL ISOLATE:
;     POSITION IN GENOME:
;       CHROMOSOME/SEGMENT:
;       MAP POSITION:
;     FEATURE:
;       NAME/KEY:
;       LOCATION:  461..728
;       IDENTIFICATION METHOD:
;       OTHER INFORMATION:
US-08-698-948-3


alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_copy_146_154 x US-08-698-948-3/rev   ..

 Align seq 1/1  to reverse of: US-08-698-948-3  from: 1  to: 268

        2 ProMetAspIrpSerPro 7
          |||||||||||||||||
      261 CCGATGGACTGGTCACCC 244

Seq name: /cgn2_6/ptodata/2/iia/1617-MB_us-09-126-945b-2.seq:US-08-105-168B-23

seq documentation block:
; Sequence 23, Application US/08105168B
; Patent No. 5589585
;  GENERAL INFORMATION:
;     APPLICANT:  MABILAT et al.
;     TITLE OF INVENTION:  DNA FRAGMENTS OF MYCOBACTERIA, AMPLIFICATION
;     TITLE OF INVENTION:  PRIMERS, HYBRIDIZATION PROBES, REAGENTS AND METHOD FOR THE
;     TITLE OF INVENTION:  MYCOBACTERIA
;     NUMBER OF SEQUENCES:  23
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Olitt & Berridge
;       STREET:  700 South Washington Street, Suite 300
;       CITY:  Alexandria,
;       STATE:  Virginia
;       ZIP:  22314
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5" DS/HD
;       COMPUTER:  IBM compatible
;       OPERATING SYSTEM:  MS DOS 3.1
;       SOFTWARE:  Wordperfect
;     CURRENT APPLICATION DATA:
;       FILING DATE:  August 12, 1993
;       APPLICATION NUMBER:  US/08/105,168B
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  FR9210094
;       FILING DATE:  August 8, 1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  William P. Berridge
;       REGISTRATION NUMBER:  30,024
;       REFERENCE/DOCKET NUMBER:  WPB 28835
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 836-6400
;       TELEFAX:  (703) 836-2787
;   INFORMATION FOR SEQ ID NO:  23:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  314 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single stranded
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:
;     ANTI SENSE:
;     ORIGINAL SOURCE:
;       ORGANISM:  Mycobacterium tuberculosis
;       STRAIN:
;       INDIVIDUAL ISOLATE:
;     POSITION IN GENOME:
```

```
;       CHROMOSOME/SEGMENT:
;       MAP POSITION:
;     FEATURE:
;       NAME/KEY:
;       LOCATION:  438-751
;       IDENTIFICATION METHOD:
;       OTHER INFORMATION:
US-08-105-168B-23


alignment_scores:
             Quality:    6.00              Length:       6
               Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-08-126-945B-2_COPY_146_154 x US-08-105-168B-23/rev

 Align seq 1/1  to reverse of: US-08-105-168B-23  from: 1  to: 414

       2 ProMetAspIrpSerPro 7
         ||||||||||||||||||
     284 CCGATGGACTGGTCACCC 267

Seq_name:  [illegible]

Seq_documentation_block:
; Sequence 23, Application US/08698948
; Patent No. 5849901
;  GENERAL INFORMATION:
;    APPLICANT:  MABILAT et al.
;    TITLE OF INVENTION:  DNA FRAGMENTS OF MYCOBACTERIA, AMPLIFICATION
;    TITLE OF INVENTION:  PRIMERS, HYBRIDIZATION PROBES, REAGENTS AND METHODS FOR THE DET
;    TITLE OF INVENTION:  MYCOBACTERIA
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Olift & Berridge
;      STREET:  700 South Washington Street, Suite 300
;      CITY:  Alexandria,
;      STATE:  Virginia
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" DS/HD
;      COMPUTER:  IBM compatible
;      OPERATING SYSTEM:  MS DOS 3.1
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/698,948
;      FILING DATE:  August 16, 1996
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/105,168
;      FILING DATE:  August 12, 1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR9210094
;      FILING DATE:  August 8, 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  William P. Berridge
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 29845A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6400
;      TELEFAX:  (703) 836-2787
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  414 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:
;    ANTI-SENSE:
;    ORIGINAL SOURCE:
;      ORGANISM:  Mycobacterium tuberculosis
;      STRAIN:
;      INDIVIDUAL ISOLATE:
;    POSITION IN GENOME:
;      CHROMOSOME/SEGMENT:
;      MAP POSITION:
;    FEATURE:
;      NAME/KEY:
;      LOCATION:  438-751
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:
US-08-698-948-23


alignment_scores:
             Quality:    6.00              Length:       6
               Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_146_154 x US-08-698-948-23/rev

 Align seq 1/1  to reverse of: US-08-698-948-23  from: 1  to: 414

       2 ProMetAspIrpSerPro 7
         ||||||||||||||||||
     284 CCGATGGACTGGTCACCC 267

Seq_name:  [illegible]

Seq_documentation_block:
; Sequence 23, Application US/08473020A
; Patent No. 5677274
;  GENERAL INFORMATION:
;    APPLICANT:  Hance, Allan J.
;    APPLICANT:  Grandchamp-Desraux, Bernard
;    APPLICANT:  Levy-Frebault, Veronique
;    APPLICANT:  Gicquel, Briqitte
;    TITLE OF INVENTION:  Nucleotide sequences of actinomycetales,
;    TITLE OF INVENTION:  applications to the synthesis or detection of nucleic
;    TITLE OF INVENTION:  acids, products of expression of such sequences and
;    TITLE OF INVENTION:  application as immunogenic compositions.
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Walter H. Dreger
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/473,020A
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/624,729
;      FILING DATE:  14-DEC-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dreger, Walter H
;      REGISTRATION NUMBER:  24190
;      REFERENCE/DOCKET NUMBER:  A54435
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  342 base pairs
;      TYPE:  nucleic acid
```

```
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US 08 473 020A 29


alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US 09 126 945B 2 COPY 146 154 x US 08 473 020A 29/rev  ..

 Align seq 1/1  to reverse of: US 08 473 020A 29  from: 1  to: 342

        2 ProMetAspTrpSerPro 7
          TTTTTTTTTTTTTTTTTT
      278 CCGATGGACTGGTCACCG 261

Seq name:  Seq. 6/pta/data/Tra/16 * Me Seq      * 5'   , 3'  1

seq documentation block:
; Sequence 41, Application US/08473020A
; Patent No. 5877273
;  GENERAL INFORMATION:
;     APPLICANT:  Rance, Allan J.
;     APPLICANT:  Grandchamp-Desraux, Bernard
;     APPLICANT:  Levy-Frebault, Veronique
;     APPLICANT:  Gicquel, Briqitte
;     TITLE OF INVENTION:  Nucleotide sequences of actinomycetales,
;     TITLE OF INVENTION:  applications to the synthesis or detection of nucleic
;     TITLE OF INVENTION:  acids, products of expression of such sequences and
;     TITLE OF INVENTION:  application as immunogenic compositions.
;     NUMBER OF SEQUENCES:  41
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Walter H. Dreger
;       STREET:  4 Embarcadero Center, Suite 3400
;       CITY:  San Francisco
;       STATE:  California
;       COUNTRY:  U.S.A.
;       ZIP:  94111
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/473,020A
;       FILING DATE:  06-JUN-1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/623,729
;       FILING DATE:  14 DEC 1990
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Dreger, Walter H
;       REGISTRATION NUMBER:  24190
;       REFERENCE/DOCKET NUMBER:  A54435
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 781 1989
;       TELEFAX:  (415) 398 4249
;   INFORMATION FOR SEQ ID NO:  41:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  342 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US 08 473 020A 41


alignment scores:
              Quality:     6.00              Length:       6
```

```
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US 09 126 945B 2 COPY 146 154 x US 08 473 020A 41/rev  ..

 Align seq 1/1  to reverse of: US 08 473 020A 41  from: 1  to: 342

        2 ProMetAspTrpSerPro 7
          TTTT TTTTTTTTTTTTT
      278 CCGATGGACTGGTCACCG 261

 Seq name:  Seq. 6/pta/data/Tra/16 * Me.seq.US on 473 020A 2

seq documentation block:
; Sequence 2, Application US/08473020A
; Patent No. 5877273
;  GENERAL INFORMATION:
;     APPLICANT:  Rance, Allan J.
;     APPLICANT:  Grandchamp-Desraux, Bernard
;     APPLICANT:  Levy-Frebault, Veronique
;     APPLICANT:  Gicquel, Briqitte
;     TITLE OF INVENTION:  Nucleotide sequences of actinomycetales,
;     TITLE OF INVENTION:  applications to the synthesis or detection of nucleic
;     TITLE OF INVENTION:  acids, products of expression of such sequences and
;     TITLE OF INVENTION:  application as immunogenic compositions.
;     NUMBER OF SEQUENCES:  41
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Walter H. Dreger
;       STREET:  4 Embarcadero Center, Suite 3400
;       CITY:  San Francisco
;       STATE:  California
;       COUNTRY:  U.S.A.
;       ZIP:  94111
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/473,020A
;       FILING DATE:  06-JUN-1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/623,729
;       FILING DATE:  14-DEC-1990
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Dreger, Walter H
;       REGISTRATION NUMBER:  24190
;       REFERENCE/DOCKET NUMBER:  A54435
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 781 1989
;       TELEFAX:  (415) 398 4249
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  343 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  both
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US 08 473 020A 2


alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US 09 126 945B 2 COPY 146 154 x US 08 473 020A 2/rev  ..

 Align seq 1/1  to reverse of: US 08 473 020A 2  from: 1  to: 343
```

```
       2 ProMetAspIrpSerPro 7
         ||||||||||||||||||
     279 CCGATGGACTGGTCACCG 262

seq_name: [illegible]

seq_documentation_block:
; Sequence 4, Application US/08473020A
; Patent No. 5877273
;  GENERAL INFORMATION:
;    APPLICANT:  Hance, Allan J.
;    APPLICANT:  Grandchamp-Desraux, Bernard
;    APPLICANT:  Levy-Frebault, Veronique
;    APPLICANT:  Gicquel, Brigitte
;    TITLE OF INVENTION:  Nucleotide sequences of actinomycetales,
;    TITLE OF INVENTION:  applications to the synthesis or detection of nucleic
;    TITLE OF INVENTION:  acids, products of expression of such sequences and
;    TITLE OF INVENTION:  application as immunogenic compositions.
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Walter H. Dreger
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/473,020A
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/623,729
;      FILING DATE:  14-DEC-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dreger, Walter H.
;      REGISTRATION NUMBER:  24190
;      REFERENCE/DOCKET NUMBER:  A54435
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  343 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-473-020A 4


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_146_154 x US-08-473-020A 4/rev

 Align seq 1/1  to reverse of: US-08-473-020A 4  from: 1  to: 343

       2 ProMetAspIrpSerPro 7
         || ||||||||||||||||
     279 CCGATGGACTGGTCACCC 262

seq_name: [illegible]

seq_documentation_block:
; Sequence 22, Application US/08105168B
```

```
; Patent No. 5589585
;  GENERAL INFORMATION:
;    APPLICANT:  MABILAT et al.
;    TITLE OF INVENTION:  DNA FRAGMENTS OF MYCOBACTERIA, AMPLIFICATION
;    TITLE OF INVENTION:  IN VITRO, HYBRIDIZATION PROBES, REAGENT AND METHOD FOR DE
;    TITLE OF INVENTION:  MYCOBACTERIA
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Oblit & Berridge
;      STREET:  700 South Washington Street, Suite 400
;      CITY:  Alexandria,
;      STATE:  Virginia
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" DS/HD
;      COMPUTER:  IBM compatible
;      OPERATING SYSTEM:  MS DOS 4.1
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      FILING DATE:  August 12, 1993
;      APPLICATION NUMBER:  US/08/105,168B
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR92 10094
;      FILING DATE:  August 8, 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  William P. Berridge
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 28835
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6400
;      TELEFAX:  (703) 836-2787
;  INFORMATION FOR SEQ ID NO:  22:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  350 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:
;    ANTI-SENSE:
;    ORIGINAL SOURCE:
;      ORGANISM:  Mycobacterium tuberculosis
;      STRAIN:
;      INDIVIDUAL ISOLATE:
;    POSITION IN GENOME:
;      CHROMOSOME/SEGMENT:
;      MAP POSITION:
;    FEATURE:
;      NAME/KEY:
;      LOCATION:  438..787
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:
US-08-105-168B-22


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_146_154 x US-08-105-168B-22/rev

 Align seq 1/1  to reverse of: US-08-105-168B-22  from: 1  to: 350

       2 ProMetAspIrpSerPro 7
         |||||||||||||||||||
     284 CCGATGGACTGGTCACCC 267

seq_name: [illegible]

seq_documentation_block:
```

```
; Sequence 22, Application US/08698948
; Patent No. 5849901
; GENERAL INFORMATION:
;    APPLICANT:  MABILAT et al.
;    TITLE OF INVENTION:  DNA FRAGMENTS OF MYCOBACTERIA, AMPLIFICATION
;    TITLE OF INVENTION:  PRIMERS, HYBRIDIZATION PROBES, REAGENTS AND METHOD FOR THE DET
;    TITLE OF INVENTION:  MYCOBACTERIA
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Oliff & Berridge
;      STREET:  700 South Washington Street, Suite 300
;      CITY:  Alexandria,
;      STATE:  Virginia
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" DS/HD
;      COMPUTER:  IBM compatible
;      OPERATING SYSTEM:  MS DOS 3.1
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/698,948
;      FILING DATE:  August 16, 1996
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/105,168
;      FILING DATE:  August 12, 1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR9210094
;      FILING DATE:  August 8, 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  William P. Berridge
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 28835A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6400
;      TELEFAX:  (703) 836-2787
;  INFORMATION FOR SEQ ID NO:  22:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  350 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:
;    ANTI-SENSE:
;    ORIGINAL SOURCE:
;      ORGANISM:  Mycobacterium tuberculosis
;      STRAIN:
;      INDIVIDUAL ISOLATE:
;    POSITION IN GENOME:
;      CHROMOSOME/SEGMENT:
;      MAP POSITION:
;    FEATURE:
;      NAME/KEY:
;      LOCATION:  438-787
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:
US-08-698-948-22

alignment_scores:
           Quality:     6.00              Length:       6
             Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_146_154 x US-08-698-948-22/rev   ..

 Align seq 1/1  to reverse of: US-08-698-948-22  from: 1  to: 350

      2 ProMetAspTrpSerPro 7
        ||||||||||||||||||
    284 CCGATGGACTGGTCACCC 267
```

```
seq_name: [illegible] seq US-08-935-450-1

seq_documentation_block:
; Sequence 1, Application US/08935450
; Patent No. 5977311
; GENERAL INFORMATION:
;  APPLICANT: Nandabalan, Krishnan
;  APPLICANT: Yang, Meijia
;  APPLICANT: Schulz, Vincent
;  TITLE OF INVENTION: 53BP2 COMPLEXES
;  FILE REFERENCE: 7934-054
;  CURRENT APPLICATION NUMBER: US/08/935,450
;  CURRENT FILING DATE: 1997-09-23
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 1554
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (757)..(3771)
US-08-935-450-1

alignment_scores:
           Quality:     6.00              Length:       6
             Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_146_154 x US-08-935-450-1   ..

 Align seq 1/1  to: US-08-935-450-1  from: 1  to: 4534

      4 AspTrpSerProSerAsn 9
        |||| |||||||||||||
   1621 GATTGGAGTCCTTCAAAT 1638
```

```
CM of: US-09-126-945B-2_COPY_146_154 to: EST:*   out format : pts

Date: Sep 14, 2001  7:47 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL frames_p2n.model -DEV xlp
 Q:/cgn2_1/356***/us-09-126-945b-2_copy_146_154.olip2n.rst.1.1/*_1.*/*.pgn_poly tmp/1 1/267
 -DB=EST -QFMT fastap -SUFFIX=olip2n.rst -QGAPOP 4.500
 -GAPEXT 0.050 -MINMATCH 0.100 -LOOPCL 0.000 -LOOPEXT 0.000
 -QGAPOP 4.500 -QGAPEXT 0.050 -XGAPOP 60.000 -XGAPEXT 60.000
 -FGAPOP 6.000 -FGAPEXT 7.000 -YGAPOP 60.000 -YGAPEXT 60.000
 -DELOP 6.000 -DELEXT 7.000 -START 1 -MATRIX olinq
 -TRANS human40.cdi -LIST=45 -DOCALIGN 200 -THR_SCORE quality
 -THR_MIN=1 -ALIGN 15 -MODE=LOCAL -OUTFMT pts  NORM ext  MINLEN 0
 -MAXLEN 2000000000 -USER US09126945B_COPY_1_4645 -NCPU 6
 -ICPU 8 -LONGLOG -NO_XLPXY -WAIT -THREADS 1

Search information block:
Query: US-09-126-945B-2 COPY 146 154
Query length: 9
Database: EST:*
Database sequences: 10228115
Database length: 431459454
Search time (sec): 9087.630000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000


score_list:
Sequence              Strd Orig    ZScore    EScore Len   ! Documentation ..
qb_est97:BG179568       +    9.00  174.48   1.4e     991  ! BG179568 602328182F1 NIH_MGC_91
qb_est45:AW415885       +    7.00  145.64   54.03    198  ! AW415885 13609 MARC 4B0V Bos ta
qb_est47:AW446425       -    7.00  143.87   67.85    261  ! AW446425 85560 MARC 1B0V Bos ta
qb_est47:AW477811       +    7.00  141.47   92.26    379  ! AW477811 16767 MARC 4B0V Bos ta
qb_est84:BF196555       +    7.00  139.50   119.78   515  ! BF196555 7m97c12.x1 NCI_CGAP_Kid
qb_est40:AV663019       +    7.00  139.12   124.83   547  ! AV663019 AV663019 Bos taurus cD
em_estpl9:BF280458      -    7.00  138.78   130.26   576  ! BF280458 CA__Fbm040E16 Soares_pi
qb_qss13:AQ545136       +    6.00  128.42   492.38   137  ! AQ545136 RPCI 11-409E1.3V RPCI
qb_qss41:AA712929            6.00  125.43   721.91   218  ! AA712929 RPCI 24 130023.1V RPCI
qb_est64:BB793657            6.00  124.23   854.79   271  ! BB793657 BB793657 RIKEN full-le
qb_est3:AA158933        +    6.00  123.73   897.69   284  ! AA158933 zo73t11.r1 Stratagene
qb_est60:BB265094       -    6.00  123.71   900.48   285  ! BB265094 BB265094 RIKEN full-le
qb_est61:BB305110       +    6.00  123.49   926.25   295  ! BB305110 BB305110 RIKEN full-le
qb_qss5:B47892          -    6.00  123.42   934.00   298  ! B47892 RPCI11-305.1V RPCI 11 Bo
qb_est5:AV258026        -    6.00  123.39   962.32   306  ! AV258026 AV258026 RIKEN full le
qb_est49:AW626906       -    6.00  122.97   990.46   320  ! AW626906 pa21b07.y1 Hawaiian Amey
qb_est106:N34586        +    6.00  122.79   1.0e+03  329  ! N34586 yx61b11.r1 Soares melanoc
qb_est51:AW757722       -    6.00  122.46   1.1e+03  346  ! AW757722 874003812.y1 C. reinhar
qb_est14:AA985309       -    6.00  122.05   1.1e+03  369  ! AA985309 am75q03.s1 Stratagene
qb_qss14:AQ590819       +    6.00  122.02   1.1e+03  371  ! AQ590819 HS_2104_A2_G11_I7C G11
qb_est107:R35697        +    6.00  122.00   1.1e+03  372  ! R35697 yq67b08.r1 Soares infant
qb_est84:BF188064       +    6.00  121.48   1.2e+03  403  ! BF188064 EST444351 potato stolo
qb_est3:AA205054        +    6.00  121.42   1.2e+03  407  ! AA205054 zq70d02.r1 Stratagene
qb_est78:BE699665       -    6.00  121.39   1.2e+03  409  ! BE699665 CM4 NN0084 1(600 195
qb_est102:BG606010      +    6.00  121.14   1.3e+03  425  ! BG606010 RH172_8D_C09.b1_A003 R
qb_est77:BE647340       -    6.00  121.13   1.3e+03  426  ! BE647340 WHE1817-1820_C05_C05Z5
qb_est109:198728        +    6.00  121.11   1.3e+03  427  ! 198728 yo61q06.r1 Soares fetal
qb_est28:AL450809       -    6.00  120.95   1.3e+03  438  ! AL450809 AL450809 Hordeum vulga
qb_est90:BF653462       -    6.00  120.95   1.3e+03  438  ! BF653462 277401 MARC 4B0V Bos t
qb_est45:AW345267       -    6.00  120.91   1.3e+03  441  ! AW345267 25238 MARC 4B0V Bos ta
qb_est12:AA801525            6.00  120.82   1.3e+03  447  ! AA801525 BL06548.5prime BL Dros
qb_est18:AI253404       -    6.00  120.78   1.3e+03  450  ! AI253404 aq14d04.x1 Stanley Fro
qb_est7:AA444027        -    6.00  120.76   1.3e+03  451  ! AA444027 zc14c11.r1 Soares ovar
qb_est71:BE197140       -    6.00  120.66   1.3e+03  458  ! BE197140 cq72t05.y1 Soares mamm
qb_est3:AA179465        +    6.00  120.63   1.3e+03  460  ! AA179465 zp45t11.r1 Stratagene
qb_est78:BE749906       +    6.00  120.63   1.3e+03  460  ! BE749906 200984 MARC 4B0V Bos t
qb_est107:R53434        +    6.00  120.56   1.3e+03  465  ! R53434 yq83b08.r1 Soares infant
qb_est107:R19656        +    6.00  120.47   1.4e+03  472  ! R19656 yq37f07.r1 Soares infant
qb_est107:R51428        +    6.00  120.29   1.4e+03  485  ! R51428 yq72e08.r1 Soares infant
qb_est78:BE750325       -    6.00  120.23   1.4e+03  490  ! BE750325 201722 MARC 4B0V Bo
qb_est16:AI161842       -    6.00  120.19   1.4e+03  493  ! AI161842 Au0f1680 Hybrid osp
qb_est90:BF652674       -    6.00  120.14   1.4e+03  497  ! BF652674 267215 MARC 4B0V Bo
qb_est2:AA089014             6.00  120.11   1.4e+03  499  ! AA089014 ml57h04.r1 Stratage
qb_est107:R59476        -    6.00  119.92   1.5e+03  514  ! R59476 yh17c09.r1 Soares int
qb_qss16:AQ784974       -    6.00  119.86   1.5e+03  519  ! AQ784974 BS_4141_A1 A10_1 T7

seq_name: qb_est97:BG179568

seq_documentation_block:
LOCUS       BG179568          991 bp    mRNA            EST       06-FEB-2001
DEFINITION  602328182F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4429414 5',
            mRNA sequence.
ACCESSION   BG179568
VERSION     BG179568.1  GI:12686271
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 991)
  AUTHORS   [illegible]
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CHTN/IDP
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10181  row: d  column: 22
            High quality sequence stop: 655.
FEATURES             Location/Qualifiers
     source          1..991
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4429414"
                     /clone_lib="NIH_MGC_91"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage resistant)"
                     /note="Stratagene prostate, Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed;
                     Average insert size 1.4 Kb. Library enriched for
                     full length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      190 a    351 c    305 g    145 t
ORIGIN


alignment_scores:
            quality:    9.00              Length:       9
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_146_154 x BG179568  .

 Align Set 1/1  to: BG179568  from: 1  to: 991

        1 AspProMetAspTrpSerProSerAsn 9
          |||||||||||||||||||||||||||
      439 GATCCCATGGACTGGAGCCCCAGCAAT 465

seq_name: qb_est45:AW415885

seq_documentation_block:
LOCUS       AW415885          198 bp    mRNA            EST       10-JUL-2000
DEFINITION  13609 MARC 4B0V Bos taurus cDNA 5', mRNA sequence.
ACCESSION   AW415885
VERSION     AW415885.1  GI:6745141
```

```
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 198)
  AUTHORS   Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
            Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
            ,W.W. and Keele,J.W.
  TITLE     Design and use of four pooled tissue normalized cDNA libraries for
            EST discovery in cattle
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4466
            Fax: 402 762 4490
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and trimmed with phred
            v0.980904.e. Vector identified by cross_match with the  minscore 20
            and  minmatch 12 options.
            PCR Primers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 11  row: N  column: 19
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/qualifiers
     source          1..198
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 4B0V"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site 1: XbaI; Site 2: XhoI;
                     Library made from pooled tissue from marrow, alveolar
                     macrophage, ovary, fetal semitendonosus muscle, and fetal
                     longissimus muscle."
BASE COUNT        51 a      52 c      47 g      48 t
ORIGIN


alignment_scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 DS_09_126_945B_2_COPY_146_154 x AW415885_...

 Align seq 1/1  to: AW415885  from: 1  to: 198

        1 AspProMetAspTrpSerPro 7
          ||||||||||||||||||||||
       56 GACCCCATGGACTGGAGCCCA 75

Seq_name: qb_est47:AW446425

Seq_documentation_block:
LOCUS       AW446425        261 bp    mRNA            EST       07-JUL-2000
DEFINITION  85560 MARC 1B0V Bos taurus cDNA 5', mRNA sequence.
ACCESSION   AW446425
VERSION     AW446425.1  GI:6988212
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 261)
  AUTHORS   Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
            Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
            ,W.W. and Keele,J.W.
  TITLE     Design and use of four pooled tissue normalized cDNA libraries for
            EST discovery in cattle
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4466
            Fax: 402 762 4490
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and trimmed with phred
            v0.980904.e. Vector identified by cross_match with the  minscore 20
            and  minmatch 12 options.
            PCR Primers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 47  row: J  column: 15
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/qualifiers
     source          1..261
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 1B0V"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site 1: XbaI; Site 2: XhoI;
                     Library made from pooled tissue from lymph node, ovary,
                     fat, hypothalamus, and pituitary."
BASE COUNT        80 a      54 c      55 g      72 t
ORIGIN


alignment_scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 DS_09_126_945B_2_COPY_146_154 x AW446425/rev_...

 Align seq 1/1  to reverse of: AW446425  from: 1  to: 261

        1 AspProMetAspTrpSerPro 7
          |||| |||||||||||||||||
       46 GATCCCATGGACTGGAGTCCA 26

Seq_name: qb_est47:AW477811

Seq_documentation_block:
LOCUS       AW477811        379 bp    mRNA            EST       10-JUL-2000
DEFINITION  16767 MARC 4B0V Bos taurus cDNA 5', mRNA sequence.
ACCESSION   AW477811
VERSION     AW477811.1  GI:7047917
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 379)
  AUTHORS   Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
            Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
            ,W.W. and Keele,J.W.
  TITLE     Design and use of four pooled tissue normalized cDNA libraries for
            EST discovery in cattle
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4466
            Fax: 402 762 4490
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and trimmed with phred
            v0.980904.e. Vector identified by cross_match with the  minscore 20
```

```
            and -mismatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 8  row: N  column: 19
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .379
                     /organism="Bos taurus"
                     /db_xref "taxon:9913"
                     /clone_lib "MARC 06V"
                     /tissue_type "pooled"
                     /lab_host="DH10B"
                     /note "Vector: pCMV SPORT; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from marrow, alveolar
                     macrophage, ovary, fetal semitendonosus muscle, and fetal
                     longissimus muscle."
BASE COUNT      117 a     91 c     82 g     89 t
ORIGIN


alignment_scores:
            Quality:    7.00             Length:      7
              Ratio:   1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_146_154 x AW477811    ..

 Align seq 1/1  to: AW477811  from: 1  to: 379

       1 AspProMetAspTrpSerPro 7
         |||||||||||||||||||||
      55 GACCCCATGGACTGGAGCCCA 75

seq_name: qb_est84:BF196555

seq_documentation_block:
LOCUS        BF196555       515 bp     mRNA             EST        03-NOV-2000
DEFINITION  7m97c12.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:3563134 3'
            similar to TR:Q14682 Q14682 KIAA0177 PROTEIN ., mRNA sequence.
ACCESSION   BF196555
VERSION     BF196555.1  GI:11084613
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 515)
  AUTHORS   NCI-NCBS CGAP http://www.ncbi.nlm.nih.gov/ncicgap
  TITLE     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL. Send email to
            info@image.llnl.gov
            High quality sequence stop: 494.
FEATURES             Location/Qualifiers
     source          1. .515
                     /organism="Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone "IMAGE:3563134"
                     /clone_lib "NCI_CGAP_Brn23"
                     /tissue_type="glioblastoma (pooled)"
                     /lab_host="DH10B"
                     /note "Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTIACGAATCTGAAGTGGGAGCGGCTGCATAITTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      135 a     103 c      95 g     179 t       2 others
ORIGIN


alignment_scores:
            Quality:    7.00             Length:      7
              Ratio:   1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_146_154 x BF196555    ..

 Align seq 1/1  to: BF196555  from: 1  to: 515

       3 MetAspTrpSerProSerAsn 9
         ||| |||||||||||||||||
     252 ATGGATTGGAGTCCCAGCAAC 252

seq_name: qb_est40:AV663019

seq_documentation_block:
LOCUS        AV663019       547 bp     mRNA             EST        25-AUG-2000
DEFINITION  AV663019 Bos taurus brain fetus Bos taurus cDNA clone E1BR034B12
            5', mRNA sequence.
ACCESSION   AV663019
VERSION     AV663019.1  GI:9922049
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos
REFERENCE   1  (bases 1 to 547)
  AUTHORS   Sugimoto,Y., Hirotsuno,S., Takasuga,A., Itoh,R., Jitohzono,A. and
            Suzuki,H.
  TITLE     bovine cDNA sequencing
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshikazu Sugimoto
            Animal Genetics Division
            Shirakawa Institute of Animal Genetics
            Odakura, Nishiso, Nishi shirakawa, Fukushima 961 8061, Japan
            Tel: 81 248-25-5641
            Fax: 81 248-25-5725
            Email: kazusugi@cocoa.ocn.ne.jp
            Single pass sequencing.
            This clone was obtained from a polyA-deleted cDNA library.
FEATURES             Location/Qualifiers
     source          1. .547
                     /organism="Bos taurus"
                     /db_xref "taxon:9913"
                     /clone "E1BR034B12"
                     /clone_lib "Bos taurus brain fetus"
                     /tissue_type "brain"
                     /dev_stage "fetus"
                     /lab_host="DH10B"
                     /note "Vector: pZLL; Site_1: SalI; Site_2: NotI; 1.0; A
                     was deleted from a NotI site"
BASE COUNT      129 a     131 c     148 g     139 t
ORIGIN


alignment_scores:
            Quality:    7.00             Length:      7
```

```
            Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-9458-2_COPY_146_154 x AV663019  ..

 Align seq 1/1  to: AV663019  from: 1  to: 547

        1 AspProMetAspIrpSerPro 7
          |||||||||||||||||||||
      357 GACCCCATGGACTGGAGCCCA 377

seq name: em-estpl9:BF280458

seq documentation block:
ID   BF280458   standard; RNA; EST; 576 BP.
XX
AC   BF280458;
XX
SV   BF280458.1
XX
DT   20-NOV-2000 (Rel. 65, Created)
DT   20-NOV-2000 (Rel. 65, Last updated, Version 1)
XX
DE   GA_Eb0040F16 Gossypium arboreum 7-10 dpa fiber library Gossypium arboreum
DE   cDNA clone GA_Eb0040F16, mRNA sequence.
XX
KW   EST.
XX
OS   Gossypium arboreum
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Malvales; Malvaceae; Gossypium.
XX
RN   [1]
RP   1-576
RA   Wing R.A., Frisch D., Yu Y., Main D., Rambo T., Simmons J., Henry  D.,
RA   Wood T.C., Leslie A., Wilkins T.A.;
RT   "An integrated analysis of the genetics, development, and evolution of the
RT   cotton fiber";
RL   Unpublished.
XX
CC   Contact: Wing RA
CC   Clemson University Genomics Institute
CC   Clemson University
CC   100 Jordan Hall, Clemson, SC 29634, USA
CC   Tel: 864 656 7288
CC   Fax: 864 656 4293
CC   Email: rwing@clemson.edu
CC   High quality sequence start  82
CC   High quality sequence stop.  298.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..576
FT                   /db_xref "taxon:29729"
FT                   /db_xref "ESTLIB:2480"
FT                   /note "Vector: pBK-CMV; Site 1: EcoRI; Site 2: XhoI"
FT                   /organism "Gossypium arboreum"
FT                   /strain "AKA"
FT                   /cultivar "8400"
FT                   /clone "GA_Eb0040F16"
FT                   /clone_lib "Gossypium arboreum 7-10 dpa fiber library"
FT                   /tissue_type "Fibers isolated from bolls harvested 7-10
FT                   dpa"
FT                   /lab_host "E. coli"
XX
SQ   Sequence 576 BP; 150 A; 138 C; 140 G; 157 T; 1 other;


alignment scores:
              Quality:    7.00              Length:      7
                Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-9458-2_COPY_146_154 x BF280458/rev  ..

 Align seq 1/1  to: reverse of: BF280458  from: 1  to: 576

        2 ProMetAspIrpSerProSer 8
          |||||||||||||||||||||
      192 CCCATGGACTGGCTCACCTGC 172

seq name: gi_gss|s:AQ535136

seq documentation block:
LOCUS       AQ535136       137 bp    DNA              GSS       18-MAY-1999
DEFINITION  RPCI-11-409E4.TV RPCI-11 Homo sapiens genomic clone RPCI-11-409E4,
            DNA sequence.
ACCESSION   AQ535136
VERSION     AQ535136.1  GI:4846826
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 137)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Marek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Shujun Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbeetigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter.dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC resources (http://bacpac.med.buffalo.edu/orderinf.htm) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end/bac_end_search.html.
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..137
                     /organism="Homo sapiens"
                     /db_xref="GDB:7656771"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11 409E4"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI-11 Human Male BAC Library"
BASE COUNT        34 a       33 c       43 g       27 t
ORIGIN


alignment scores:
              Quality:    6.00              Length:      6
                Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-9458-2_COPY_146_154 x AQ535136  ..

 Align seq 1/1  to: AQ535136  from: 1  to: 137

        1 AspProMetAspIrpSer 6
          ||||||||||||||||||
       57 GATCCCATGGACTGGAGC 74
```

```
seq_name: gb_gss41:AZ712929

seq_documentation_block:
LOCUS       AZ712929       218 bp    DNA             GSS       24-JAN-2001
DEFINITION  RPCI-24-130D23.TV RPCI-24 Mus musculus genomic clone RPCI-24-130D23
            , DNA sequence.
ACCESSION   AZ712929
VERSION     AZ712929.1  GI:12447153
KEYWORDS    GSS.
SOURCE      house mouse
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 218)
  AUTHORS   Zhao,S., Nierman,W., Mulok,I., Shetty,S., Akiyama,J., Tsegaye,G.,
            Geer,K., Krol,M., Shvartsbeyn,A., Schreiber,E., Russell,D., de
            Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-130D23.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.chori.org). Clones may be purchased from BACPAC
            Resources (http://www.bacpac.chori.org/forpurchasinginfo.htm). BAC end
            pages: http://www.tigr.org/tdb/tgi/mgi/mse/bac_ends/ [see
            Plate: 130  row: D  column: 23
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..218
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-130D23"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamHI sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT       64 a     46 c     59 g     49 t
ORIGIN


alignment_scores:
            Quality:    6.00            Length:       6
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_146_154 x AZ712929/rev  ..

 Align seq 1/1  to reverse of AZ712929  from  1  to  218

        4 MetAspIrpSerProSer 8
          ||||||||||||||||||
      119 ATGGACTGGAGCCCTTCA 102

seq_name: gb_est64:BB393657

seq_documentation_block:
LOCUS       BB393657       271 bp    mRNA            EST       13-JUL-2000
DEFINITION  BB393657 RIKEN full-length enriched, 0 day neonate cerebellum Mus
            musculus cDNA clone C230084N18 3' similar to X64810 M.musculus mRNA
            for 11 like protein, mRNA sequence.
ACCESSION   BB393657
VERSION     BB393657.1  GI:9116635
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 271)
  AUTHORS   Kono,H., Aizawa,K., Arakawa,S., Akiyama,J., Arakawa,I., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,I., Itoh,M.,
            Iwama,M., Kodeki,F., Kogiwa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,R., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shimagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Taqawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,I., Yamanaka,I.
            , Yano,K., Yasunishi,A., Yokota,I., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Kono,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: yoshi@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
            Itoh,M., Kitamura,S., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Ishwara,Y., Carninci,P., Shibata,Y., Ohsawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
            Automated filtration based high throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            High efficiency full length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1..271
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="C230084N18"
                     /clone_lib "RIKEN full length enriched, 0 day neonate
                     cerebellum"
                     /tissue_type "cerebellum"
                     /dev_stage "0 day neonate"
                     /lab_host="DH10B"
                     /note "Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGCATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3']. cDNA was
                     prepared by using trehalose thermo activated reverse
                     transcriptase and subsequently enriched for full length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot= 20.0 and subtraction to Rot= 479.. 2nd
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAAATTAATCCCCCCCCCCCCCC
```

```
                    3'). cDNA was cleavzed with XhoI and BamHI. Vector: a
                    modified pBluescript KS(-) after bulk excision from Lambda
                    FLC 1."
BASE COUNT        81 a      72 c      61 g      57 t
ORIGIN


alignment scores:
            Quality:     6.00              Length:       6
              Ratio:    1.000               Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2 (COPY 146-154 x BG936652/qy)    ..

 Align seq 1/1   to reverse of: BG936652   from: 1   to: 271

        3 MetAspTrpSerProSer 8
          ||||||||||||||||||
       46 ATGGACTGGTCTCCTCT 19

Seq name: qb|est4:AA158933

Seq documentation block:
LOCUS         AA158933        284 bp     mRNA             EST        08-NOV-1997
DEFINITION  zc73111.r1 Stratagene pancreas (#937208) Homo sapiens cDNA clone
            IMAGE:592557 5' similar to qb|D14748 EUKARYOTIC INITIATION FACTOR
            4A I (HUMAN);, mRNA sequence.
ACCESSION     AA158933
VERSION       AA158933.1  GI:1733744
KEYWORDS      EST.
SOURCE        human.
  ORGANISM    Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1  (bases 1 to 284)
  AUTHORS     Hillier,L., Allen,M., Bowles,L., Duduque,T., Geisel,C., Jost,S.,
              Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
              ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
              White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE       Wash0 NCI human EST Project
  JOURNAL     Unpublished (1997)
COMMENT       Contact: Wilson RK
              Washington University School of Medicine
              4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
              Tel: 314 286 1800
              Fax: 314 286 1810
              Email: est@watson.wustl.edu
              This clone is available royalty free through LLNL ; contact the
              IMAGE Consortium (info@image.llnl.gov) for further information.
              Insert Length: 1498  Std Error: 0.00
              Seq primer:  28MI3 rev2 from Amersham
              High quality sequence stop: 155.
FEATURES             Location/Qualifiers
     source          1..284
                     /organism="Homo sapiens"
                     /db_xref="GDB:4623943"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:592557"
                     /clone_lib="Stratagene pancreas (#937208)"
                     /lab_host="SOLR cells (kanamycin resistant)"
                     /note="organ: pancreas; Vector: pBluescript SK ; Site 1:
                     EcoRI; Site 2: XhoI; Cloned unidirectionally.  Primer:
                     oligo dT. Pancreatic adenocarcinoma cell line. Average
                     insert size: 1.0 Kb; Bal YAP X8 Vector;  5' adaptor
                     sequence: 5' GAATTCGGCACGAG 3'  3' adaptor sequence: 5'
                     CTCGAGTTTTTTTTTTTTTTT 3'"
BASE COUNT        64 a      66 c      72 g      69 t      13 others
ORIGIN


alignment scores:
            Quality:     6.00              Length:       6
              Ratio:    1.000               Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2 (COPY 146-154 x AA158933)    ..

 Align seq 1/1   to: AA158933   from: 1   to: 284

        2 ProMetAspTrpSerPro 7
          |||| ||||||||||||
       31 CCGATCGACTCGAGCCCG 48

Seq name: qb|est4:BG265094

Seq documentation block:
LOCUS         BG265094        285 bp     mRNA             EST        07-JUL-2000
DEFINITION  BG265094 RIKEN full length enriched, 10 days neonate cortex Mus
            musculus cDNA clone A830018P14 3' similar to X94410 M.musculus mRNA
            for 11 like protein, mRNA sequence.
ACCESSION     BG265094
VERSION       BG265094.1  GI:18961594
KEYWORDS      EST.
SOURCE        house mouse.
  ORGANISM    Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE     1  (bases 1 to 285)
  AUTHORS     Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
              ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
              Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,I., Itoh,M.,
              Iwata,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kiyuchi,N.,
              Kiyosawa,H., Kojima,Y., Kondo,S., Koyn,S., Kurihara,C., Kusakabe,M.
              , Matsuyama,I., Miki,E., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
              , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
              ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
              , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomitaka,N., Toya
              ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,I., Yamanaka,I.
              , Yano,R., Yasunishi,A., Yokota,I., Yoshida,K., Yoshiki,A., Yoshino
              ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE       RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL     Unpublished (2000)
COMMENT       Contact: Yoshihide Hayashizaki
              Genome Exploration Research Group, Life Science Tsukuba Center,
              Genome Science Laboratory
              The Institute of Physical and Chemical Research (RIKEN), Genomic
              Sciences Center
              3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
              Tel: +81 298-36-9013
              Fax: +81 298-36-9098
              Email: genome-res@rtc.riken.go.jp,
              URL:http://genome.rtc.riken.go.jp/
              Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
              ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
              Thermostabilization and thermoactivation of thermolabile enzymes by
              trehalose and its application for the synthesis of full length
              cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
              Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
              Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
              ,Y. and Hayashizaki,Y.
              Automated filtration based high-throughput plasmid preparation
              system. Genome Res. 9 (5), 463-470 (1999)
              Carninci,P. and Hayashizaki,Y.
              High efficiency full length cDNA cloning. Methods Enzymol. 303,
              19-44 (1999)
              Please visit our web site (http://genome.rtc.riken.go.jp) for
              further details.
FEATURES             Location/Qualifiers
     source          1..285
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="A830018P14"
                     /clone_lib="RIKEN full-length enriched, 10 days neonate
                     cortex"
```

```
                     /tissue_type "cortex"
                     /dev_stage="10 days neonate"
                     /lab_host "DH10B"
                     /note "Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot  20.0 and subtraction to Rot  459.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGAATTCTCGAGTTAATTAAATTAATTTTTTTTTTTT
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(-) after bulk excision from Lambda
                     FLC I."
BASE COUNT        84 a      66 c      70 g      65 t
ORIGIN


alignment_scores:
             Quality:     6.00              Length:       6
               Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_146_154 x BB265094/rev    .

 Align seq 1/1  to reverse of. BB265094  from: 1  to: 285

        3 MetAspTrpSerProSer 8
          |||||||||||||||||||
       48 ATGGACTGGTCTCCGTCT 31

seq_name: gb_est61:BB305110

seq_documentation_block:
LOCUS        BB305110       295 bp     mRNA             EST       10-JUL-2000
DEFINITION   BB305110 RIKEN full-length enriched, adult male corpora
             quadrigemina Mus musculus cDNA clone B2301117G09 3', mRNA sequence.
ACCESSION    BB305110
VERSION      BB305110.1  GI:9005815
KEYWORDS     EST.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 295)
  AUTHORS    Kohno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,I., Carninci
             ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
             Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,I., Itoh,M.,
             Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
             Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
             , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Ota,H., Okazaki,Y.
             , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
             ,Y., Shiqemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
             , Suzuki,H., Suzuki,H., Tarui,A., Tashiro,F., Tominaga,N., Toya
             ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,I., Yamanaka,I.
             , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
             ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE      RIKEN Mouse ESTs (Kohno,H., et al.)
  JOURNAL    Unpublished (2000)
COMMENT      Contact: Yoshihide Hayashizaki
             Genome Exploration Research Group, Life Science Tsukuba Center,
             Genome Science Laboratory
             The Institute of Physical and Chemical Research (RIKEN), Genomic
             Sciences Center
             3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
             Tel: +81-298-36-9013
             Fax: +81-298-36-9098
             Email: genome-res@ttc.riken.go.jp,
             ***** http://genome.gsc.riken.go.jp/
             Site_1: SalI  Site_2: BamHI  cDNA library was prepared and
             sequenced in Mouse Genome Encyclopedia Project of Genome
             Exploration Research Group in Riken Genomic Sciences Center and
             Genome Science Laboratory in RIKEN. Division of Experimental
             Animal Research in Riken contributed to prepare mouse tissues.
             1st strand cDNA was primed with a primer [5'
             GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
             prepared by using trehalose thermo-activated reverse
             transcriptase and subsequently enriched for full-length by
             cap-trapper. cDNA went through one round of normalization
             to Rot  20.0 and subtraction to Rot  459.0. Second
             strand cDNA was prepared with the primer adapter of
             sequence [5' GAGAGAGAGAATTCTCGAGTTAATTAAATTAATTTTTTTTTTTT
             3']. cDNA was cleaved with XhoI and BamHI. Vector: a
             modified pBluescript KS(-) after bulk excision from Lambda
             FLC I."
BASE COUNT        93 a      57 c      61 g      84 t
ORIGIN


alignment_scores:
             Quality:     6.00              Length:       6
               Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_146_154 x BB305110    .

 Align seq 1/1  to: BB305110  from: 1  to: 295

        4 AspTrpSerProSerAsn 9
          |||||||||||||||||||
      236 GACTGGTCTCCGTCTAGC 253

seq_name: gb_gss85:B47892

seq_documentation_block:
LOCUS        B47892         298 bp     DNA              GSS       08-AUG-1999
DEFINITION   RPCI11-325.1V RPCI-11 Homo sapiens genomic clone RPCI-11 325, DNA
             sequence.
ACCESSION    B47892
VERSION      B47892.1  GI:2600129
KEYWORDS     GSS.
```

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 298)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
            ,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences for Sequence Ready Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are derived from the Human BAC library, RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources. (http://bacpac.chori.org/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/cgi-bin/BAC/BAC_End_search.cgi
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..298
                     /organism="Homo sapiens"
                     /db_xref="GDB:758008"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11 30"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT       92 a      61 c      76 g      69 t
ORIGIN


alignment scores:
            Quality:     6.00            Length:       6
              Ratio:    1.000              Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2 COPY 146-154 x B47802/rev

 Align seq 1/1  to reverse of B47802  from 1  to 298

        4 AspTrpSerProSerAsn 9
          ||||||||||||||||||
       97 GACTGGTCTCCAICCAAC 80

seq name: gb:est5:AV258026

seq documentation block:
LOCUS       AV258026        309 bp    mRNA            EST       04 NOV 1999
DEFINITION  AV258026 RIKEN full length enriched, adult male testis (DH10B) Mus
            musculus cDNA clone 4922501G06 3' similar to X64410 M.musculus mRNA
            for L1 like protein, mRNA sequence.
ACCESSION   AV258026
VERSION     AV258026.1  GI:6245485
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 309)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
            Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
            Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kanagawa,T., Kai
            ,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kosakabe,M.,
            Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Ooba,H., Okazaki,Y.,
            Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata
            ,Y., Shiraganeto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suroki,H.,
            Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N., Tsunoda,Y.,
            Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokoyama,T.,
            Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al. 1999)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp,
            http://genome.rtc.riken.go.jp/
            Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tsujita,T., Yoneda,Y.,
            Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki
            ,Y.
            "Transcriptional sequencing: A method for DNA sequencing using RNA
            [ ]-exonuclease", Nat. Acad. Sci. U.S.A. 96 (2), 9450-9455 (1999)
            Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
            "Automated filtration-based high-throughput plasmid preparation
            system", Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            "High efficiency full-length cDNA cloning", Methods Enzymol. 303,
            19-44 (1999)
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1..309
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="4922501G06"
                     /clone_lib="RIKEN full length enriched, adult male testis
                     (DH10B)"
                     /sex="male"
                     /tissue_type="testis"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in RIKEN
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN, Division of Experimental Animal Research in RIKEN
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3']. cDNA was
                     prepared by using trehalose thermo activated reverse
                     transcriptase and subsequently enriched for full length by
                     cap trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAAATCCCCCCCCCCCC 3']. cDNA
                     was cloned into the XhoI and BamHI sites. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."
BASE COUNT       87 a      68 c      86 g      68 t
ORIGIN


alignment scores:
            Quality:     6.00            Length:       6
              Ratio:    1.000              Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2 COPY 146-154 x AV258026/rev
```

Alignment 1/1 to reverse of: AV25402F from 1 to 369

```
  1 MetAspIrpSerProSer 8
    ||||||||:|||||||||
 72 AtGGACtGGtCtCCCtCt 55
```

```
OM of: US-09-126-945B-2_COPY_165_173 to: GenEmbl:*   out_format : pts

Date: Sep 14, 2001  9:47 PM

About: Results were produced by the GenCore software, version 4.5
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
  MODEL frames_p2n.model -DEV-xlp
  -Q "/cgn2_1/tmp/US09_126_945B_2_COPY_165_173_1:[0-1]/seq_pre_start_1.1247
  -DB GenEmbl -QFMT fastap -SUFFIX olip2n.rge -GAPOP=4.500
  GAPEXT 0.050 -MINMATCH 0.100  100PCL=0.000 -100PEXT 0.000
  -QGAPOP=4.500  QGAPEXT=0.050 -XGAPOP=60.000  XGAPEXT 60.000
  -FGAPOP=6.000  FGAPEXT=7.000 -YGAPOP=60.000  YGAPEXT 60.000
  -DELOP=6.000 -DELEXT 7.000 -START=1 -MATRIX clipo
   TRANS human40.cdi -LIST 45 -DOCALIGN=200 -IBR_SCORE=quality
  -THR_MIN=1 -ALIGN 15 -MODE LOCAL -OUTFMT pts  NORM ext  MINLEN 0
  -MAXLEN=2000000000 -USER US09126945_COGN1_1_7594 -NCPU=6
  -ICPU=4 -LONGLOG -NO_XLPXY -WAIT  THREADS 1

Search information block:
Query: US-09-126-945B-2_COPY_165_173
Query length: 9
Database: GenEmbl:*
Database sequences: 1344157
Database length: -856069004
Search time (sec): 6503.700000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000


score_list:
Sequence              Strd Orig    ZScore    EScore Len  ! Documentation ..
qb_pat1:AX048148        +    9.00  160.00  1.1e-13  1894  ! Ax048148 Sequence 1 from Patent
qb_pr4:AF071538         +    9.00  160.00  1.1e-13  1894  ! AF071538 Homo sapiens Ets tran
qb_pr1:AB031549         +    9.00  159.99  1.1e-13  1898  ! AB031549 Homo sapiens PSE mRNA
qb_pat1:AX048159        +    9.00  157.02  1.6e-13  3317  ! AX048159 Sequence 12 from Pate
qb_pr6:AL157372         +    9.00  136.70  22.33    150660  ! AL157372 Human DNA sequence
qb_ba1:AF071024         -    7.00  122.94  130.46   2165  ! AF071024 Streptomyces chrysomu
qb_pl1:AC004786         -    7.00  105.20  1.3e-08  60597  ! AC004786 Arabidopsis thaliana
qb_htq6:AC018743_2      -    7.00  102.03  1.9e-08  110000  ! Continuation (3 of 4) of AC
qb_htq15:AC069227       +    7.00   99.67  2.5e-08  171318  ! AC069227 Homo sapiens chromo
qb_htq15:AC073214       +    7.00   99.59  2.6e-08  174114  ! AC073214 Homo sapiens chromo
qb_htq15:AC069138       -    7.00   99.28  2.7e-08  184346  ! AC069138 Homo sapiens chromo
qb_htq17:AC079952       -    7.00   99.19  2.7e-08  187692  ! AC079952 Homo sapiens chromo
qb_htq10:AC026108       +    7.00   99.08  2.8e-08  191476  ! AC026108 Homo sapiens chromo
qb_htq13:AC064874       +    7.00   99.05  2.8e-08  192487  ! AC064874 Homo sapiens chromo
qb_htq15:AC073157       +    7.00   97.88  3.2e-08  240068  ! AC073157 Homo sapiens chromo
qb_pl3:AY019015         +    6.00  116.26  307.39   250  ! AY019015 Oryza sativa microsate
qb_in4:IVY09492         -    6.00  114.50  395.27   348  ! S.cerevisiae chromosome XRNA
qb_sts2:G29697          -    6.00  113.45  446.56   432  ! G29697 human STS SHGC-34875, se
qb_sts2:G62401          -    6.00  112.94  470.26   466  ! G62401 RH61A14/SP6 Human Chromo
qb_sts2:G54187          -    6.00  112.55  494.77   502  ! G54187 RH61A9/SP6 Human Chromos
qb_vi1:AF297278         -    6.00  112.53  496.12   501  ! AF297278 Bovine viral diarrhea
qb_pr5:AF309471S2       -    6.00  112.28  512.13   528  ! AF309472 Homo sapiens SOX6 (SOX
qb_pl2:AF325042         +    6.00  111.17  590.22   650  ! AF325042 Arabidopsis thaliana A
qb_sts1:CNS00XQ0        -    6.00  111.05  599.49   665  ! AL420070 17 end of clone AY0AAC
qb_ov:OCIS01            -    6.00  110.96  606.24   676  ! X16621 Catostomus commersoni mR
qb_ba1:AF001833         -    6.00  110.88  612.96   687  ! AF001833 Bacillus amyloliquefac
qb_ov:OCIS02            -    6.00  110.60  644.71   723  ! X16622 Catostomus commersoni mR
qb_pr4:AF018155         +    6.00  110.57  647.71   728  ! AF018155 Macaca fascicularis cl
qb_pr4:AF018153         +    6.00  110.55  649.50   731  ! AF018153 Cercopithecus aethiops
qb_pr4:AF018154         +    6.00  110.55  649.50   731  ! AF018154 Macaca fascicularis cl
qb_vi1:AF318306         -    6.00  110.43  649.02   747  ! AF318306 Dendrolimus punctatus
qb_pr9:HSLRPSS25        -    6.00  110.35  655.53   758  ! AF058421 Homo sapiens lipoprote
qb_pl2:AIPSBWGEN        +    6.00  109.68  715.11   861  ! X99769 A.thaliana pstW gene, 8/
qb_vi2:RVJ225220        -    6.00  109.49  742.10   909  ! AJ225220 RV Palmate newt parti
qb_htq15:AC069935       -    6.00  109.45  745.44   915  ! AC069935 Giardia intestinalis c
qb_htq18:AC086612       +    6.00  109.24  756.52   945  ! AC086612 Giardia intestinalis c
qb_htq13:AC062626       +    6.00  108.93  787.17   991  ! AC062626 Giardia intestinalis c
qb_htq13:AC047258       -    6.00  108.83  796.91  1009  ! AC047258 Giardia intestinalis
qb_htq13:AC056978       -    6.00  108.83  797.19  1010  ! AC056978 Giardia intestinalis
qb_sts1:CNS09SZS        -    6.00  108.59  822.07  1056  ! AL412736 17 end of clone AW
qb_sts1:CNS06QKJ        -    6.00  108.44  838.99  1088  ! AL411029 13 end of clone AW
qb_pl4:SPD97399         -    6.00  108.40  842.68  1095  ! D97399 Schizosaccharomyces
qb_htq13:AC048105       -    6.00  108.33  850.02  1109  ! AC048105 Giardia intestinal
qb_htq13:AC057095       -    6.00  108.11  874.45  1156  ! AC057095 Giardia intestinal
qb_vi2:YM0244058        -    6.00  108.00  886.80  1180  ! AJ244058 Yaba monkey virus p

seq_name: qb_pat1:AX048148

seq_documentation_block:
LOCUS       AX048148      1894 bp    DNA             PAT       15-DEC-2000
DEFINITION  Sequence 1 from Patent WO0070092.
ACCESSION   AX048148
VERSION     AX048148.1  GI:11876967
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1894)
  AUTHORS   Kaufmann,J. and Xin,H.
  TITLE     Expression of ets-domain proteins in cancer
  JOURNAL   Patent: WO 0070092-A 1 23-NOV-2000;
            Chiron Corporation (US)
FEATURES             Location/Qualifiers
     source          1..1894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             416..1423
                     /note="Human JKEIS."
                     /codon_start=1
                     /protein_id="AAL19002.1"
                     /db_xref="GI:11876968"
                     /translation="MGSASPGLSSVSPSHLELPPDIVSPTGLEKAAAGAVGLEEKLWS
                     RSETAIEQGLSAEYLSYFDMLYFLDASWAAKAEGASEEEFEELEEQFVTIFKAPA
                     GSLDLVPGALELEEESLEQVQSMVVGEVLKPLEIAFKLLNLTAPPMDWSPSNVCRWLL
                     WIEHQYRLPPMGKAFQELAGKELCAMSEEQFRQRSPLGGDVLRAHLDIWKSAAWMKER
                     TSFGAIHYCASTSEESWTDSEVDSCSGQFIHLWYELKELLKNRHSYGKFISWLNKEK
                     GIFSIESDSAQVAELWGIRKNREAMNYDKLSRSIRQYYEKYCTEKPDISQKLVYQFVHP
                     I"
BASE COUNT      368 a    654 c    571 g    302 t
ORIGIN


alignment_scores:
            Quality:    9.00              Length:      9
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_165_173 x AX048148  ..

 Align seq 1/1  to: AX048148  from: 1  to: 1894

       1 GluLysArgLeuProProMetGlyLys 9
         ||||||||||||||||||||||||||
     908 GAATACCGGCTGCCCCCCATGGGCAAG 934

seq_name: qb_pr4:AF071538

seq_documentation_block:
LOCUS       AF071538      1894 bp    mRNA            PRI       10-JAN-2000
DEFINITION  Homo sapiens Ets transcription factor ERF (ERF) mRNA, complete
            cds.
ACCESSION   AF071538
VERSION     AF071538.1  GI:4007417
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1894)
  AUTHORS   Oettgen,P., Finger,E., Sun,Z., Akbarali,Y., Thamrongsak,U.,
```

```
                Boltax,J., Grall,F., Dube,A., Weiss,A., Brown,L., Quinn,G., Kas,K.,
                Endress,G., Kunsch,C. and Libermann,T.A.
  TITLE         PDEF, a novel prostate epithelium specific ets transcription
                factor, interacts with the androgen receptor and activates
                prostate specific antigen gene expression
  JOURNAL       J. Biol. Chem. 275 (2), 1216-1225 (2000)
  MEDLINE       20092894
REFERENCE       2  (bases 1 to 1894)
  AUTHORS       Libermann,T.A., Oettgen,P., Kunsch,C. and Finget,E.
  TITLE         Direct Submission
  JOURNAL       Submitted (10-JUN-1998) Medicine, Beth Israel Deaconess Medical
                Center, 330 Brookline Ave., Boston, MA 02215, USA
FEATURES             Location/Qualifiers
     source          1..1894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6p21.3"
                     /tissue_type="prostate"
     gene            1..1894
                     /gene="PDEF"
     CDS             416..1423
                     /gene="PDEF"
                     /note="prostate epithelium specific"
                     /codon_start=1
                     /product="Ets transcription factor PDEF"
                     /protein_id="AAC97206.1"
                     /db_xref="GI:4007418"
                     /translation="MGSASPGLSSVSPSHLLLPPDTVSRTGLEKAAAGAVGLERRDWS
                     PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAPGASSREEPPEEPEQCPVTDSQAPA
                     GSLDLVPGGLTLEEHSLEQVQSMVVCEVLKDTEIACKLLNITADPMDWSPSNVQKWLL
                     WTEHQYKLEPMGKAEQLALGKLLCAMSELQTFQRSFLGGFVEHAHLFCWKEAAWRLER
                     TSPGATHYCASTSEESWTDSEVDSSCSGQPTHEWQFTEKLLERHEYGRTCEWLNERK
                     GTFKTLDSAQVARLWGTRKNRPAMNYDKLSRGTRQYYKKGTTRKPDISQRLVYQFVHP
                     I"
BASE COUNT       368 a    654 c    571 g    302 t
ORIGIN


alignment scores:
            Quality:     9.00              Length:       9
              Ratio:    1.000               Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2_COPY_165_173 x AF071538   ..

 Align seq 1/1  to: AF071538  from: 1  to: 1894

        1 GlnTyrArgLeuProProMetGlyLys 9
          |||||||||||||||||||||||||
      908 CAATACCGGCTGCCCCCCATGGGCAAG 934

seq_name: qb_pri:AB031549

seq_documentation_block:
LOCUS        AB031549      1898 bp    mRNA            PRI       20-JAN-2000
DEFINITION   Homo sapiens PSE mRNA for prostate ets, complete cds.
ACCESSION    AB031549
VERSION      AB031549.1  GI:6721497
KEYWORDS     prostate ets.
SOURCE       Homo sapiens adult male prostate epithelium prostate cancer
             cell line:PC-3 cDNA to mRNA.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (sites)
  AUTHORS    Yamada,N., Tamai,Y., Miyamoto,H. and Nozaki,M.
  TITLE      Cloning and expression of the mouse Pse gene encoding a novel Ets
             family member
  JOURNAL    Gene 241 (2), 267-274 (2000)
  MEDLINE    20147504
REFERENCE    2  (bases 1 to 1898)
  AUTHORS    Nozaki,M.
  TITLE      Direct Submission
  JOURNAL    Submitted (25-AUG-1999) to the DDBJ/EMBL/GenBank databases. Masami
             Nozaki, Osaka University, Institute for Microbial Diseases; 3-1,
             Yamadaoka, Suita, Osaka 565-0871, Japan
             (E-mail:mnozaki@biken.osaka-u.ac.jp, Tel:81-6-879-8338,
             Fax:81-6-879-8339)
FEATURES             Location/Qualifiers
     source          1..1898
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="PC-3"
                     /cell_type="prostate cancer"
                     /chromosome="6"
                     /dev_stage="adult"
                     /map="6p21.2-21.3"
                     /sex="male"
                     /tissue_type="prostate epithelium"
                     /note="vector:pBC18"
     gene            406..1413
                     /gene="PSE"
     CDS             406..1413
                     /gene="PSE"
                     /codon_start=1
                     /product="prostate ets"
                     /protein_id="BAA89544.1"
                     /db_xref="GI:6721498"
                     /translation="MGSASPGLSSVSPSHLLLPPDTVSRTGLEKAAAGAVGLERRDWS
                     PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAPGASSREEPPEEPEQCPVTDSQAPA
                     GSLDLVPGGLTLEEHSLEQVQSMVVGEVLKDTEIACKLLNITADPMDWSPSNVQKWLL
                     WTEHQYKLEPMGKAEQLALGKLLCAMSELQTFQRSFLGGFVEHAHLFCWKEAAWRLER
                     TSPGATHYCASTSEESWTDSEVDSSCSGQPTHEWQFTEKLLERHEYGRTCEWLNERK
                     GTFKTLDSAQVARLWGTRKNRPAMNYDKLSRGTRQYYKKGTTRKPDISQRLVYQFVHP
                     I"
BASE COUNT       380 a    651 c    567 g    400 t
ORIGIN


alignment scores:
            Quality:     9.00              Length:       9
              Ratio:    1.000               Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2_COPY_165_173 x AB031549   ..

 Align seq 1/1  to: AB031549  from: 1  to: 1898

        1 GlnTyrArgLeuProProMetGlyLys 9
          |||||||||||||||||||||||||
      898 CAATACCGGCTGCCCCCCATGGGCAAG 924

seq_name: qb_pat1:AX048159

seq_documentation_block:
LOCUS        AX048159      3317 bp    DNA             PAT       15-DEC-2000
DEFINITION   Sequence 12 from Patent WO0070092.
ACCESSION    AX048159
VERSION      AX048159.1  GI:11876975
KEYWORDS     .
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 3317)
  AUTHORS    Kaufmann,J. and Xin,H.
  TITLE      Expression of ets-domain proteins in cancer
  JOURNAL    Patent: WO 0070092-A 12 23-NOV-2000;
             Chiron Corporation (US)
FEATURES             Location/Qualifiers
     source          1..3317
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

```
                     /note "GSEF encoding sequence with promoter (Figs. 2A 2B)"
     promoter        1. .1380
                     /note "Promoter"
     CDS             1815. .2822
                     /note "GSEF coding sequence"
                     /codon_start 1
                     /protein_id "CAC19894.1"
                     /db_xref "GI:11876976"
                     /translation="MGRASFLSAVVAFCMGLIFSRLIVSKIVPKAAASAVGLFKPLWS
                     PSPPATPEQGLSAFYLSYFIMLYPEISSWAAAKAPGASSGLLPPEEPEVGPVHNSQAPA
                     GSLDLVPGGLTLFPHSLFQVQSMVVGFVLKDLEFAGKLLNLIALPMDWSPSNVGKWLL
                     WIFHGYPLPPMGFASQFEACFLGAMSLGFMQFSICGVVLHASLDIWFSAAWMEEP
                     TSPGAIHYCASTSEESWIDSFVDSSCSGQPIHLWQFLKPLLLKLPHSYCSFLEWLNKPK
                     GIFKIEDSAQVARLWGIRFNNFAMNYPKQSFLQVLPLLIIEFIGLSKLVYGPVIH
                     1"
BASE COUNT       710 a    1026 c     970 g     611 t
ORIGIN


alignment_scores:
              Quality:     9.00              Length:       9
                Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_165_173 x AX048159   ..

 Align seq 1/1  to: AX048159  from: 1  to: 3317

       1 GluTyrArgLeuProProMetGlyLys 9
         |||||||||||||||||||||||||||
    2307 CAATACCGGCTGCCCCCCATGGGCAAG 2333

seq_name: gb_pr6:AL157372

seq_documentation_block:
LOCUS        AL157372    150660 bp    DNA              PRI       29-NOV-2000
DEFINITION   Human DNA sequence from clone RP11-375E1 on chromosome 6, complete
             sequence.
ACCESSION    AL157372
VERSION      AL157372.18  GI:11493236
KEYWORDS     HTG.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 150660)
  AUTHORS    Skuce,C.
  TITLE      Direct Submission
  JOURNAL    Submitted (29-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
             CB10 1SA, UK. Email enquiries: humquery@sanger.ac.uk Clone
             requests: clonerequest@sanger.ac.uk
COMMENT      On Nov 30, 2000 this sequence version replaced gi:11342772.
             During sequence assembly data is compared from overlapping clones.
             Where differences are found these are annotated as variations
             together with a note of the overlapping clone name. Note that the
             variation annotation may not be found in the sequence submission
             corresponding to the overlapping clone, as we submit sequences with
             only a small overlap as described above.
             This sequence has been finished according to sequence map criteria
             as follows. An attempt is made to resolve all sequencing problems,
             such as compressions and repeats, but not necessarily within known
             annotated repeat sequence elements. Where the sequence is
             ambiguous, there is an annotation using the 'unsure' feature key.
             The following abbreviations are used to associate primary accession
             numbers given in the feature table with their source databases:
             Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP. Information
             on the WORMPEP database can be found at
             http://www.sanger.ac.uk/Projects/C_elegans/wormpep. This sequence
             was generated from part of bacterial clone contigs of human
             chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
             Group. Further information can be found at
             http://www.sanger.ac.uk/HGP/Chr6.
             RP11-375E1 is from the library RPCI-11,2 constructed at the Roswell
             Park Cancer Institute by the group of Pieter de Jong. For further
             [illegible] see http://bacpac.chori.org/.
             VECTOR: pBACe3.6
             IMPORTANT: This sequence is not the entire insert of clone
             RP11-375E1 it may be shorter because we sequence overlapping
             sections only once, except for a 100 base overlap.
             The true left end of clone RP11-375E1 is at 1 in this sequence. The
             true left end of clone RP1-22 is at 150561 in this sequence.
             The true right end of clone RP11-302L10 is at 113044 in this
             sequence.
FEATURES             Location/Qualifiers
     source          1. .150660
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-375E1"
                     /clone_lib="RPCI-11,2"
BASE COUNT     45436 a   38887 c   38740 g   27564 t
ORIGIN


alignment_scores:
              Quality:     9.00              Length:       9
                Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_165_173 x AL157372/rev   ..

 Align seq 1/1  to reverse of: AL157372  from: 1  to: 150660

       1 GluTyrArgLeuProProMetGlyLys 9
         |||||||||||||||||||||||||||
  126410 CAATACCGGCTGCCCCCCATGGGCAAG 126384

seq_name: gb_bal:AF071024

seq_documentation_block:
LOCUS        AF071024     2165 bp    DNA              BCT       22-JAN-1999
DEFINITION   Streptomyces chrysomallus DNA replication initiator protein (dnaA)
             gene, complete cds.
ACCESSION    AF071024
VERSION      AF071024.1  GI:4165888
KEYWORDS     .
SOURCE       Streptomyces chrysomallus.
  ORGANISM   Streptomyces chrysomallus
             Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
             Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE    1  (bases 1 to 2165)
  AUTHORS    Majka,J., Jakimowicz,D., Messer,W., Schrempf,H., Lisowski,M. and
             Zakrzewska-Czerwinska,J.
  TITLE      Interactions of the Streptomyces lividans initiator protein DnaA
             with its target
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 2165)
  AUTHORS    Jakimowicz,D. and Zakrzewska-Czerwinska,J.
  TITLE      Direct Submission
  JOURNAL    Submitted (05-JUN-1998) Microbiology, Ludwik Hirszfeld Institute of
             Immunology and Experimental Therapy, Weigla 12, Wroclaw 53-114,
             Poland
REFERENCE    3  (bases 1 to 2165)
  AUTHORS    Jakimowicz,D. and Zakrzewska-Czerwinska,J.
  TITLE      Direct Submission
  JOURNAL    Submitted (27-OCT-1998) Microbiology, Ludwik Hirszfeld Institute of
             Immunology and Experimental Therapy, Weigla 12, Wroclaw 53-114,
             Poland
  REMARK     Sequence update by submitter
REFERENCE    4  (bases 1 to 2165)
  AUTHORS    Jakimowicz,D. and Zakrzewska-Czerwinska,J.
  TITLE      Direct Submission
  JOURNAL    Submitted (21-JAN-1999) Microbiology, Ludwik Hirszfeld Institute of
             Immunology and Experimental Therapy, Weigla 12, Wroclaw 53-114,
```

```
            Poland
  REMARK    Sequence update by submitter
COMMENT     on Jan 22, 1999 this sequence version replaced gi:3800784.
FEATURES             Location/Qualifiers
     source          1..2165
                     /organism="Streptomyces chrysomallus"
                     /strain="ATCC11523"
                     /db_xref="taxon:1899"
                     /db_xref="ATCC:11523"
     gene            291..2165
                     /gene="dnaA"
     CDS             291..2165
                     /gene="dnaA"
                     /note="DnaA"
                     /codon_start=1
                     /transl_table=11
                     /product="DNA replication initiator protein"
                     /protein_id="AAC60077.1"
                     /db_xref="GI:4800785"
                     /translation="MAEVIALTAAAVWGLVFVGLLGSQLTEPFEPGWIEGVFIAV
                     ADTALLAVPNEWGEEVLGGEAPLLGELGEEFHSTTLVDEGAGEEPGPIAEEM
                     HGSHQSQQGHKYFAQQRDDAPRGDAYDGYGHRPSDDGMPTRKPAYPDYQQQREEPGAW
                     PRTQEDLSWQQPRHGGYQDREQPSQEPYRESESYREKENEQYREQAPEQWRQPYGIGR
                     PQQPQHDYRSQPPEHQGYEQQRPDRQDQQQPRQGGHGPGRTGGSVPGPMGAQPAPAP
                     GPGEPHARLNPKYLEDFEVIGASNGFAHAAAVAVAEAPAKAYNPLFIYGESGLGKTHL
                     LHAIGHYAKSLYPGTKVKYVSSEKFNEFTNSIRDNKAKDFKRKYRDVDILLVDDIQF
                     LAGKEQTQEFFHTFNTLHNANKQIVISSDRPPKQLVTLEDRLRNRFEWGLITDVQPP
                     ELETRIAILRKKAVQEQLNAPPEVLEFIASRISRNIRELEGALIRVTAFASLNKQPVD
                     LGLIEIVLKDLIPGGEESAPETIAPATMAATAIDYPGLTVDDLQGSSRTRVLVIARQIA
                     MYLCRELTDLSLPKIGAQFGGRDHTTVMHAIRKIEALMAERETYNQVITELTNKIKNG
                     "
BASE COUNT       479 a    796 c    755 g    285 t
ORIGIN


alignment scores:
            Quality:     7.00            Length:       7
              Ratio:    1.000              Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2-COPY-165_173 x AF071024/rev  ..

 Align seq 1/1  to reverse of: AF071024  from: 1  to: 2165

        4 ArgGluProProMetGlyLys 9
          |||||||||||||||||||||
       50 CGCCTGCCTCCGATGGGCAAA 30

seq_name: gb_pll:AC004786

seq_documentation block:
LOCUS       AC004786     60597 bp    DNA             PLN       05-APR-2000
DEFINITION  Arabidopsis thaliana chromosome II section 130 of 255 of the
            complete sequence. Sequence from clones T20K9.
ACCESSION   AC004786 AE002093
VERSION     AC004786.2  GI:6598442
KEYWORDS    HTG.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 60597)
  AUTHORS   Lin,X., Kaul,S., Rounsley,S.D., Shea,T.P., Benito,M. I., Town,C.D.,
            Fujii,C.Y., Mason,T.M., Bowman,C.L., Barnstead,M.E.,
            Feldblyum,T.V., Buell,C.R., Ketchum,K.A., Lee,J.J., Ronning,C.M.,
            Koo,H., Moffat,K.S., Cronin,L.A., Shen,M., Van Aken,S.E., Umayam,L.,
            Tallon,L.J., Gill,J.E., Adams,M.D., Carrera,A.J., Creasy,T.H.,
            Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D.,
            Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and
            Venter,J.C.
  TITLE     Sequence and analysis of chromosome 2 of the plant Arabidopsis
            thaliana
  JOURNAL   Nature 402 (6763), 761-768 (1999)
  MEDLINE   20093182
   PUBMED   10617197
REFERENCE   2  (bases 1 to 60597)
  AUTHORS   Lin,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
COMMENT     on Dec 17, 1999 this sequence version replaced gi:3445196.
            The sequence and annotation of chromosome 2 were merged from those
            of the individual clones on this chromosome after removing
            overlaps. For detailed information, please see the TIGR web site
            (http://www.tigr.org/tdb/at/at.html).

            Genes were identified by a combination of three methods: Gene
            prediction programs including GRAIL
            (http://avalon.epm.ornl.gov/genlab_suite.html), GeneFinder (Phil Green,
            University of Washington), Genscan (Chris Burge,
            http://genes.mit.edu/GENSCAN.html), and BestHitGene
            (http://www.tigr.org/cgi-bin/BestHitGene/BestHitGene.cgi) searches of the
            complete sequence against a peptide database and plant EST
            databases at TIGR, and manual curations based on those analyses.
            Annotated genes are named to indicate the level of evidence for
            their annotation. Genes with similarity to other proteins are named
            after the database hits. Genes without significant peptide
            similarity but with EST similarity are named as 'unknown' proteins.
            Genes without protein or EST similarity, that are predicted by two
            or more gene prediction programs over most of their length are
            annotated as 'hypothetical' proteins. Genes encoding tRNAs are
            predicted by tRNAscan-SE (Sean Eddy,
            http://www.genetics.wustl.edu/eddy/tRNAscan-SE/). Simple repeats were
            identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/RepeatMasker.html). Genes are
            numbered from the top to bottom of the chromosome.

            We thank the CSHL/Washu/ABI consortium for sequencing BAC clones
            F6F23, F5J6, F17A5, and T13L16, the ESSA group for sequencing clone
            F13O4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards
            and Satoshi Tabata for helpful assistance. In addition, we would
            like to thank the TIGR Bioinformatics Department, especially Lixin
            Zhou, Razi Khalak, Michael E. Hosney, Lily Fu, Feng Liang, Jeremy
            Peterson, Michael Holmes, and Delwood Richardson for software and
            database support.

            This work was supported by the National Science Foundation,
            Department of Energy and the US Department of Agriculture.

            Address all correspondence to:  at@tigr.org.
FEATURES             Location/Qualifiers
     source          1..60597
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="II"
     misc_feature    1..60597
                     /note="Sequence from clone T20K9"
     mRNA            complement(join(584..984,1156..1245,1340..1786))
                     /gene="At2g22800"
     gene            complement(584..1786)
                     /gene="At2g22800"
                     /note="T20K9.1; identical to GB:U09441"
     CDS             complement(join(651..984,1156..1245,1340..1770))
                     /gene="At2g22800"
                     /codon_start=1
                     /product="homeodomain transcription factor (HAT9)"
                     /protein_id="AAC32427.1"
                     /db_xref="GI:3445197"
                     /translation="MGPDPEPNTGLVIGLGPSPTPNNYNSTTRQSSVYKLEPSSTTCL
                     SGSPSVIVVTGADQLCRQTSSHSGVSSFSSGRVVKKEKDQGEESPEEFEMIEPVISDY
                     HLETSSICARLEGCELPQQSALEELSTEDEICNPEQKQVIAKQLNLKPRQVEVWFQN
                     KKARIKLKQTEVIPELKKCPLLAPNLKQQKETQELKTLKLQFETMHMPASTLIK
                     CPSPPPPPPPPPPNGQGGGGSGATAVIVDPSTAPGAFSTSSKPHFFNPFTNPSAAC"
```

```
     repeat_region   complement(744..801)
                     /rpt_family "(CCG)n"
     repeat_region   995..1064
                     /rpt_family "(TAAAA)n"
     repeat_region   complement(2483..2545)
                     /rpt_family "(TAAA)n"
     repeat_region   complement(3469..3560)
                     /rpt_family "(TA)n"
     repeat_region   8319..8431
                     /rpt_family "(TA)n"
     repeat_region   11379..11486
                     /rpt_family "(TA)n"
     repeat_region   11966..12017
                     /rpt_family "(CAA)n"
     mRNA            complement(join(<13494..14478,14567..14727,14795..14926,
                     15029..>15175))
                     /gene="At2g22810"
     gene            complement(<13494..>15175)
                     /gene "At2g22810"
                     /note="T20K9.2; identical to CB:U23481"
     CDS             complement(join(13494..14478,14567..14727,14795..14926,
                     15029..15175))
                     /gene="At2g22810"
                     /codon_start=1
                     /product "[illegible]
                     (ACS4)"
                     /protein_id "AAC32428.1"
                     /db_xref="GI:3445198"
                     /translation "MV[illegible]
                     LAFNQ[illegible]
                     RGNRVSFDSNNLVLTAGATSANETLMFCLADPGDAFLLPTPYYPGFDRDLKWRTGVEI
                     VPIQSSSTNGFRITKLA[illegible]EEAYFQAKKLD[illegible]NVKGILITNPSNPLGTTTTQTELNILFD
                     FITKNKNIHLVSDEIYSGTVFNSSEF[illegible]SVME[illegible]LKNNQLENTDVLNRVHIV[illegible]SLSK[illegible]
                     LPGFRVGAIYSNDKDVISAATKMSSFGLVSSQTQYLLSSLLSDKKFTKNYLRENQKRL
                     KNRQRKLV[illegible]GLEKAG[illegible]KC[illegible]KSNAGLFCWVDMRPLLRSKTFEAEMDLWKKIVYEVKLNI
                     SPGSSCHCEEPGWFRVCFANMIDETLKLALKRLKMLVDDENSSRRCQKSKSERLNGSR
                     KKTMSNVSNWVFRLSFHDREAEER"
     repeat_region   complement(15408..15456)
                     /rpt_family="POLY_A"
     mRNA            complement(<16934..>17128)
                     /gene "At2g22820"
     gene            complement(<16934..>17128)
                     /gene "At2g22820"
                     /note="T20K9.3; predicted by genscan"
     CDS             complement(16934..17128)
                     /gene="At2g22820"
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="AAC32429.1"
                     /db_xref="GI:3445199"
                     /translation "MK[illegible]
                     DVGQNIRKVGTANYRKIFGR"
     repeat_region   18967..19020
                     /rpt_family "POLY_A"
     mRNA            complement(join(<19650..19748,19826..20104,20180..20843,
                     20912..21086,21174..21285,21545..>21973))
                     /gene "At2g22830"
     gene            complement(<19650..>21973)
                     /gene="At2g22830"
                     /note="T20K9.4"
     CDS             complement(join(19650..19748,19826..20104,20180..20843,
                     20912..21086,21174..21285,21545..21973))
                     /gene="At2g22830"
                     /codon_start=1
                     /product="putative squalene epoxidase"
                     /protein_id "AAC32430.1"
                     /db_xref "GI:3445200"
                     /translation="MKPFVIRNLPRRQS[illegible]
                     IPPWKATAAQ[illegible]LKISAVNSTVMMKPAKIALPQFIASL[illegible]LLLY[illegible]LRRS[illegible]NKNKKNRG
                     LVVSQNDTVSKNLETEVDSGTDVIIVGAGVAGSALAHTLGKEGRRVHVIERDFSEQ[illegible]R
                     IVGELLQPGGYLKL[illegible]VKKIDAQ[illegible]VLGYVLFKD[illegible]KHIKLAYPLETF[illegible]SDVAG
                     RSFHNGRFVQRMREKAL[illegible]SNV[illegible]IVESLLEEHG[illegible]IKGVRYRTKEGNEFRSFAPL
                     TIVCDGCFSNLRRSLCKPKVDVPSTFVGLVLENCELPFANHGHVVLGDPSPILMYPIS
                     SSEV[illegible]LVDV[illegible]GQKLP[illegible]ANG[illegible]MAKYLKTKVAPQVP[illegible]KVREAF[illegible]AVEK[illegible]NIRIMPNR
                     SM[illegible]A[illegible]GALLLG[illegible]AFNMRHPLTGGGMIVALADIVV[illegible]
                     YIESFYTLRKKPVASTINTLADALYKVFLASSDEARTEMREACFDYLSLGGVFSSGPVA
                     [illegible]NP[illegible]VLHFFAVAIYAV[illegible]EAE
                     GVRQMFFPRTIPAIYRAPP"
     repeat_region   21415..21524
                     /rpt_family "(TAAAA)n"
     repeat_region   complement(24404..24453)
                     /rpt_family "(GA)n"
     mRNA            join(<24544..24871,25035..25247,25493..25865,
                     26166..>26844)
                     /gene "At2g22840"
     gene            <24544..>26844
                     /gene "At2g22840"
                     /note "T20K9.5"
     CDS             join(24544..24871,25035..25247,25493..25865,26166..26844)
                     /gene "At2g22840"
                     /note "unknown protein"
                     /codon_start 1
                     /protein_id "AAC32431.1"
                     /db_xref "GI:3445201"
                     /translation "M[illegible]GVEV[illegible]NRQ[illegible]SSK
                     ISEISIDGESSSPASAKTISRRQ[illegible]IPLLESPLLNDPRKQ[illegible]MISFSSASCESLVSPY
                     LQYCRNSGYGLGGMMNTSNMHGNLLIGVKGPFSLTQWAELEQQALIYKYIIANVPVPS
                     [illegible]NM[illegible]RRI[illegible]KEWR[illegible]AVP
                     [illegible]KY[illegible]RHINPGRHRSRKPVEGQNG[illegible]NTNAAAAASAAAAS[illegible]AAAVSKAAAGISAVAMR
                     GSDNNNSLAAAVG[illegible]QHHPNNQSTDSLANRVQNSRGASVPPAIMNLQSK[illegible]HFKQSNNP
                     FFF[illegible]SSDS[illegible]NPSHKQASYAISSKGFGSYLDPGNQAKHAGNHNNVLSWPEEIKSDW
                     T[illegible]MAPS[illegible]VQ[illegible]KLALSPLRL[illegible]EE[illegible]AIHM[illegible]VNI[illegible]KINNWIPIS
                     W[illegible]NNSM[illegible]VENSIINSPKKE[illegible]
                     CDGKDPLGPTILMNISATAPSL"
     repeat_region   complement(25686..25738)
                     /rpt_family "(CAG)n"

alignment_scores:
            Quality:     7.00             Length:       7
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_165_173 x AC004786 ..

 Align seq 1/1 to: AC004786 from: 1 to: 60597

       3 ArgLeuProProMetGlyLys 9
         |||| ||||||||||||||||
   24464 CGATTACCTCCTATGGGAAAA 24484

seq name: gb_htg6:AC018743_2

seq_documentation_block:
WPCOMMENT
  Sequence split into 4 fragments  LOCUS AC018743 Accession AC018743
     Fragment Name        Begin      End
     AC018743_0               1      110000
     AC018743_1          100001      210000
     AC018743_2          200001      310000
     AC018743_3          300001      351979
Continuation (3 of 4) of AC018743 from base 200001 (AC018743 Homo sapiens chromosome


alignment_scores:
            Quality:     7.00             Length:       7
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_165_173 x AC018743_2/rev ..

 Align seq 1/1 to reverse of: AC018743_2 from: 1 to: 110000

       3 ArgLeuProProMetGlyLys 9
         |||||||||||||||||||||
```

```
   43909 AGACTTCCTCCCATGGGTAAA 43889

Seq_name: gb_htg1:AC069227

seq_documentation_block:
LOCUS       AC069227    171418 bp    DNA              HTG       14 APR 2001
DEFINITION  Homo sapiens chromosome 12 clone RP11 478G16, WORKING DRAFT
            SEQUENCE, 4 unordered pieces.
ACCESSION   AC069227
VERSION     AC069227.18  GI:13493014
KEYWORDS    HTG; HIGS_PHASE1; HIGS_DRAFT; HIGS_EDITOR; HIGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 171418)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Alfosman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunga,H.C., Are,T.R., Banks,T., Barbaria,T.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,I.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,P., Davila,M.L., Davis,C.,
            Davy Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Dena,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,F., Ferrando,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,L.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,F., Harris,C., Harris,F., Hart,M., Hawes,A.,
            Hernandez,J., Hernandez,Y., Hodgson,A., Holder,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,R., Jiao,Y., Johnson,R., Jolivet,S.,
            Jondah,S., Karlsson,E., Kelly,S., Khan,Z., Kirst,L., Kovar,J.,
            Kovar,C., Kratovic,J., Kuresh,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Liebtanje,C., Liou,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., Mcleod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Ogoh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rivers,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,C., Shen,H., Shewbhari,N.,
            Sisson,I., Soderqren,E., Sonaike,T., Sparks,A., Stanley,R.,
            Stone,R., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,I., Teltrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wieczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.
            and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 171418)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (22 MAY 2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Apr 1, 2001 this sequence version replaced gi:13430174.
                        Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu
                Contact: hgsc.help@bcm.tmc.edu
                        Project Information
                Center project name: RREF
                Center clone name: RP11 478G16
                        Summary Statistics
                Sequencing vector: M13; L08821
                Chemistry: Dye primer Bodipy; 45% of reads
                Chemistry: Dye terminator Big Dye; 55% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 174314 bases at least Q40
                Consensus quality: 177883 bases at least Q30
                Consensus quality: 179546 bases at least Q20
                Estimated insert size: 168416; sum of contigs estimation
                quality coverage: 8x in Q20 bases; agarose fp estimation
                quality coverage: 6.9x in Q20 bases; sum of contigs estimation
                .
              * NOTE: Estimated insert size may differ from sequence length
              *   (see http://www.hgsc.bcm.tmc.edu/faq-draft-data.html).
              * NOTE: This is a 'working draft' sequence. It currently
              * consists of 4 contigs. The true order of the pieces
              * is not known and their order in this sequence record is
              * arbitrary. Gaps between the contigs are represented as
              * runs of N, but the exact sizes of the gaps are unknown.
              * This record will be updated with the finished sequence
              * as soon as it is available and the accession number will
              * be preserved.
              *       1    94964: contig of 94964 bp in length
              *   94965   95064: gap of unknown length
              *   95065  168532: contig of 73468 bp in length
              *  168533  168632: gap of unknown length
              *  168633  169825: contig of 1193 bp in length
              *  169826  169925: gap of unknown length
              *  169926  171418: contig of 1493 bp in length.
FEATURES             Location/Qualifiers
     source          1..171418
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11 478G16"
BASE COUNT     [illegible]
ORIGIN


alignment_scores:
            quality:     7.00              length:       7
              Ratio:  1.000                Gaps:         0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 GB:39 126.95*B 2_*PY_165_173 x AC069227/rev  ..

 Align seq 1/2  to reverse of: AC069227  from: 1  to: 171418

        8 ArgSerProHroMetGlyLys 9
          |||||||||||||||||||
   119836 AGACTTCCTCCCATGGGTAAA 119816

Seq_name: gb_htg1:AC073214

seq_documentation_block:
LOCUS       AC073214    174114 bp    DNA              HTG       12 JUN 2000
DEFINITION  Homo sapiens chromosome 15 clone RP11 494F2, WORKING DRAFT
            SEQUENCE, 33 unordered pieces.
ACCESSION   AC073214
VERSION     AC073214.1  GI:8440014
KEYWORDS    HTG; HIGS_PHASE1; HIGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 174114)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of H.sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 174114)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (10 JUN 2000) Genome Sequencing Center, Washington
```

```
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT
             --  ------------ Genome Center ------ ---  .
            * NOTE: this is a "working draft" sequence. It currently
            * consists of 33 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1      1268: contig of 1268 bp in length
            *    1269      1368: gap of unknown length
            *    1369      2435: contig of 1067 bp in length
            *    2436      2535: gap of unknown length
            *    2536      3932: contig of 1397 bp in length
            *    3933      4032: gap of unknown length
            *    4033      5072: contig of 1040 bp in length
            *    5073      5172: gap of unknown length
            *    5173      6502: contig of 1330 bp in length
            *    6503      6602: gap of unknown length
            *    6603      8370: contig of 1768 bp in length
            *    8371      8470: gap of unknown length
            *    8471     10138: contig of 1668 bp in length
            *   10139     10238: gap of unknown length
            *   10239     12507: contig of 2269 bp in length
            *   12508     12607: gap of unknown length
            *   12608     14490: contig of 1883 bp in length
            *   14491     14590: gap of unknown length
            *   14591     16752: contig of 2162 bp in length
            *   16753     16852: gap of unknown length
            *   16853     20317: contig of 3465 bp in length
            *   20318     20417: gap of unknown length
            *   20418     24108: contig of 3691 bp in length
            *   24109     24208: gap of unknown length
            *   24209     27209: contig of 3001 bp in length
            *   27210     27309: gap of unknown length
            *   27310     31292: contig of 3983 bp in length
            *   31293     31392: gap of unknown length
            *   31393     34758: contig of 3366 bp in length
            *   34759     34858: gap of unknown length
            *   34859     39704: contig of 4846 bp in length
            *   39705     39804: gap of unknown length
            *   39805     43000: contig of 3196 bp in length
            *   43001     43100: gap of unknown length
            *   43101     47739: contig of 4639 bp in length
            *   47740     47839: gap of unknown length
            *   47840     53040: contig of 5201 bp in length
            *   53041     53140: gap of unknown length
            *   53141     58569: contig of 5429 bp in length
            *   58570     58669: gap of unknown length
            *   58670     63993: contig of 5324 bp in length
            *   63994     64093: gap of unknown length
            *   64094     71342: contig of 7249 bp in length
            *   71343     71442: gap of unknown length
            *   71443     79045: contig of 7603 bp in length
            *   79046     79145: gap of unknown length
            *   79146     86058: contig of 6913 bp in length
            *   86059     86158: gap of unknown length
            *   86159     93434: contig of 7276 bp in length
            *   93435     93534: gap of unknown length
            *   93535    101753: contig of 8219 bp in length
            *  101754    101853: gap of unknown length
            *  101854    110615: contig of 8762 bp in length
            *  110616    110715: gap of unknown length
            *  110716    119256: contig of 8541 bp in length
            *  119257    119356: gap of unknown length
            *  119357    127879: contig of 8523 bp in length
            *  127880    127979: gap of unknown length
            *  127980    137448: contig of 9469 bp in length
            *  137449    137548: gap of unknown length
            *  137549    148758: contig of 11210 bp in length
            *  148759    148858: gap of unknown length
            *  148859    160579: contig of 11721 bp in length
            *  160580    160679: gap of unknown length
            *  160680    174114: contig of 13435 bp in length.
FEATURES             Location/Qualifiers
     source          1..174114
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /clone="RPCI 494F2"
BASE COUNT    52550 a  35771 c  34202 g  48474 t    4217 others
ORIGIN


alignment_scores:
              quality:     7.60              Length:       7
                Ratio:    1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_165_173 x AC073214    ..

 Align Seq 1/1  to: AC073214  from: 1  to: 174114

        2 TyrArgGlnProProMetGly 8
          |||| |.|||||||||||||||
   89510 TACCGGCTCCCCCCCATGGGA 89530

seq_name: gb_htg15:AC069138

seq_documentation_block:
LOCUS        AC069138    184346 bp    DNA              HTG       07-JUL-2000
DEFINITION   Homo sapiens chromosome 15 clone RPCI 81-H22, WORKING DRAFT
             SEQUENCE, 21 unordered pieces.
ACCESSION    AC069138
VERSION      AC069138.2  GI:8469088
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 184346)
  AUTHORS    Waterston,R.H.
  TITLE      The sequence of Homo sapiens clone
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 184346)
  AUTHORS    Waterston,R.H.
  TITLE      Direct Submission
  JOURNAL    Submitted (18-MAY-2000) Genome Sequencing Center, Washington
             University School of Medicine, 4444 Forest Park Parkway, St. Louis,
             MO 63108, USA
COMMENT      On Jun 12, 2000 this sequence version replaced gi:7924614.

             --  ---- -------- Genome Center - ---- -  --
             Center: Washington University Genome Sequencing Center
             Center code: WUGSC
             Web  ---- ---- /---- ---- ----/-----/-----/-----/index.html
                  -- ------ Project Information --  -  ---
             Center project name: H_NB081-0H22
                  -  ----- Summary Statistics  -  -
             Sequencing vector: M13; 100%
             Sequencing vector: plasmid; 0%
             Chemistry: dye-primer ET; 100% of reads
             Chemistry: dye-terminator Big Dye; 0% of reads
             Assembly program: Phrap; version 0.990319
             Consensus quality: 174752 bases at least Q40
             Consensus quality: 176814 bases at least Q30
             Consensus quality: 178206 bases at least Q20
             Insert size: 187000; agarose-fp
             Insert size: 182346; sum-of-contigs
             Quality coverage: 4.54 in Q20 bases; agarose fp
             Quality coverage: 4.68 in Q20 bases; sum of contigs
                  ------------------------------ ---  .
```

```
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 21 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1235: contig of 1235 bp in length
            *    1236     1335: gap of unknown length
            *    1336     3195: contig of 1860 bp in length
            *    3196     3295: gap of unknown length
            *    3296     4772: contig of 1477 bp in length
            *    4773     4872: gap of unknown length
            *    4873     7155: contig of 2283 bp in length
            *    7156     7255: gap of unknown length
            *    7256    10123: contig of 2868 bp in length
            *   10124    10223: gap of unknown length
            *   10224    14138: contig of 3915 bp in length
            *   14139    14238: gap of unknown length
            *   14239    18620: contig of 4382 bp in length
            *   18621    18720: gap of unknown length
            *   18721    24723: contig of 6003 bp in length
            *   24724    24823: gap of unknown length
            *   24824    31241: contig of 6418 bp in length
            *   31242    31341: gap of unknown length
            *   31342    37657: contig of 6316 bp in length
            *   37658    37757: gap of unknown length
            *   37758    45066: contig of 7309 bp in length
            *   45067    45166: gap of unknown length
            *   45167    53743: contig of 8577 bp in length
            *   53744    53843: gap of unknown length
            *   53844    63278: contig of 9435 bp in length
            *   63279    63378: gap of unknown length
            *   63379    72957: contig of 9579 bp in length
            *   72958    73057: gap of unknown length
            *   73058    83275: contig of 10218 bp in length
            *   83276    83375: gap of unknown length
            *   83376    96686: contig of 13311 bp in length
            *   96687    96786: gap of unknown length
            *   96787   111411: contig of 14625 bp in length
            *  111412   111411: gap of unknown length
            *  111412   125894: contig of 14483 bp in length
            *  125895   125994: gap of unknown length
            *  125995   140658: contig of 14664 bp in length
            *  140659   140758: gap of unknown length
            *  140759   160507: contig of 19749 bp in length
            *  160508   160607: gap of unknown length
            *  160608   184346: contig of 23739 bp in length.
FEATURES             Location/Qualifiers
     source          1..184346
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /clone="RP11 810H22"
     misc_feature    1..1235
                     /note="assembly_name:Contig12"
     misc_feature    1336..3195
                     /note="assembly_name:Contig14"
     misc_feature    3296..4772
                     /note="assembly_name:Contig15"
     misc_feature    4873..7155
                     /note="assembly_name:Contig16"
     misc_feature    7256..10123
                     /note="assembly_name:Contig17"
     misc_feature    10224..14138
                     /note="assembly_name:Contig18"
     misc_feature    14239..18620
                     /note="assembly_name:Contig19"
     misc_feature    18721..24723
                     /note="assembly_name:Contig20"
     misc_feature    24824..31241
                     /note="assembly_name:Contig21"
     misc_feature    31342..37657
                     /note="assembly_name:Contig22"
     misc_feature    37758..45066
                     /note="assembly_name:Contig23"
     misc_feature    45167..53743
                     /note="assembly_name:Contig24"
     misc_feature    53844..63278
                     /note="assembly_name:Contig25"
     misc_feature    63379..72957
                     /note="assembly_name:Contig26"
     misc_feature    73058..83275
                     /note="assembly_name:Contig27"
     misc_feature    83376..96686
                     /note="assembly_name:Contig28"
     misc_feature    96787..111411
                     /note="assembly_name:Contig29"
     misc_feature    111412..125894
                     /note="assembly_name:Contig30"
     misc_feature    125995..140658
                     /note="assembly_name:Contig31"
     misc_feature    140759..160507
                     /note="assembly_name:Contig32
                     clone end:17
                     vector_side:right"
     misc_feature    160608..184346
                     /note="assembly_name:Contig33"
BASE COUNT    55036 a  36111 c  36782 g  54413 t    2004 others
ORIGIN


alignment scores:
                Quality:     7.00              Length:       7
                  Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_165_173 x AC069138/rev   ..

 Align seq 1/1  to reverse of: AC069138  from: 1  to: 184346

        2 TyrArgGlnProProMetGly 8
          ||||||||||||||||||||
    24582 ATACGCCAGCCCCCGATGGGA 24562

seq_name: gb:htgl7:AC079952

seq_documentation_block:
LOCUS       AC079952    187692 bp    DNA              HTG       12-MAR-2001
DEFINITION  Homo sapiens chromosome 12q clone RP11 467A19, WORKING DRAFT
            SEQUENCE, 9 unordered pieces.
ACCESSION   AC079952
VERSION     AC079952.13  GI:13270497
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 187692)
  AUTHORS   Mungall,D.M., Adams,C., Adio-oduola,B., Ali osman,F.R., Allen,C.,
            Ainsbrooks,S.L., Amaratunga,H.C., Are,J.K., Banks,I., Barboria,J.,
            Beatson,L., Bimage,K., Blankenburg,K., Bonnin,J., Boone,K.J.,
            Bowles,S., Brieva,M., Brown,E., Brown,M., Bryand,N.P., Bubuy,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,I.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,K.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyne,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy Carroll,L., Dedorich,D.A., Delaney,K.R., Delgado,O.,
            Dean,A.L., Diaz,Y., Dinh,H.H., Douthwaite,K.J., Draper,B.,
            Duran Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Ellaj,C., Escotto,M., Falls,F., Ferraquto,D., Flaqa,N., Ford,L.,
            Foster,J., Frantz,F., Gabisi,A., Gao,J., Garcia,A., Garner,I.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
```

```
            Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Kervah,J.,
            Kovar,C., Kratovic,T., Kuroshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Orajuniye,N., Oviedo,R., Pace,A., Payton,B.,
            Perry,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojohekan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Soderqren,E., Somaiko,P., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlinqton,S., Williams,G., Williamson,A., Wlozyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.
            and Gibbs,R
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 187692)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-SEP-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Mar 10, 2001 this sequence version replaced gi:12957563.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HCFI
                Center clone name: RP11-467A19
            -------------- Summary Statistics
                Sequencing vector: M13; L08821
                Chemistry: Dye primer BodIpy; 22% of reads
                Chemistry: Dye terminator Big Dye; 78% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 185341 bases at least Q40
                Consensus quality: 190086 bases at least Q30
                Consensus quality: 192149 bases at least Q20
                Estimated insert size: 184871; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; sum-of-contigs estimation
                Quality coverage: 5.7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/Genome_draft_info.html)
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    83966: contig of 83966 bp in length
            *   83967    84066: gap of unknown length
            *   84067   121490: contig of 37424 bp in length
            *  121491   121590: gap of unknown length
            *  121591   143973: contig of 22383 bp in length
            *  143974   144073: gap of unknown length
            *  144074   163473: contig of 19400 bp in length
            *  163474   163573: gap of unknown length
            *  163574   179618: contig of 16145 bp in length
            *  179619   179718: gap of unknown length
            *  179719   183758: contig of 4040 bp in length
            *  183759   183858: gap of unknown length
            *  183859   185021: contig of 1163 bp in length
            *  185022   185121: gap of unknown length
            *  185122   186529: contig of 1408 bp in length
            *  186530   186629: gap of unknown length
            *  186630   187692: contig of 1063 bp in length.
FEATURES                Location/Qualifiers
     source             1..187692
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="12q"
                        /clone="RP11-467A19"
BASE COUNT     55595 a   47878 c   48457 g   54953 t     809 others
ORIGIN


alignment_scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 HS-09-126-945b-2_COPY_165_173 x AC079952/rev   ..

 Align seq 1/1  to reverse of: AC079952  from: 1  to: 187692

       3 ArgLeuProProMetGlyLys 9
         ||||||||||||||||||||
  163758 AGACTTCCTCCCATGGGTAAA 163738

seq_name: gb_htg10:AC026108

seq_documentation_block:
LOCUS       AC026108   191476 bp    DNA             HTG       15-MAR-2001
DEFINITION  Homo sapiens chromosome 12 clone RP11-438H24, WORKING DRAFT
            SEQUENCE, 6 unordered pieces.
ACCESSION   AC026108
VERSION     AC026108.16  GI:13346519
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 191476)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunqe,H.C., Are,J.R., Banks,I., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,P., Beuck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Bobay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Caracos,S.R., Chacko,J., Chavez,D., Chen,C., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoules,C., Cleveland,C.C., Cog,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Dicks,A.L., Ding,Y., Dinh,H.H., Dorthwriter,S.L., Draper,H.,
            Duqan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falis,T., Ferraquto,D., Flaqq,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Kervah,J.,
            Kovar,C., Kratovic,T., Kuroshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Orajuniye,N., Oviedo,R., Pace,A., Payton,B.,
```

```
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,C., Shen,H., Shooshtari,N.,
            Sisson,I., Soderquron,E., Sonaike,T., Sparks,A., Stanley,R.,
            Stone,H., Sutton,A., Svatek,A., Taboe,P., Tamerisa,A., Tamerisa,K.
            Tang,H., Tansey,J., Taylor,C., Taylor,I., Telfrod,R., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,C., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.
            and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 191476)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAR-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Mar 15, 2001 this sequence version replaced gi:12658647.
                  -- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@hgsc.bcm.edu
                  -- Project Information
            Center project name: HAMN
            Center clone name: RP11-438H24
                  -- Summary Statistics
            Sequencing vector: M13; L08821
            Chemistry: Dye primer Bodipy; 14% of reads
            Chemistry: Dye terminator Big Dye; 86% of reads
            Assembly program: Phrap; version 0.990429
            Consensus quality: 187492 bases at least Q40
            Consensus quality: 191480 bases at least Q30
            Consensus quality: 193591 bases at least Q20
            Estimated insert size: 190587; sum of-contigs estimation
            Quality coverage: 0x in Q20 bases; agarose-fp estimation
            Quality coverage: 5.7x in Q20 bases; sum of contigs estimation
            ---------- ------
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/Default.aspx?pageid=draft_data.htm).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 6 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *       1    93663: contig of 93663 bp in length
          *   93664    93763: gap of unknown length
          *   93764   141996: contig of 48233 bp in length
          *  141997   142096: gap of unknown length
          *  142097   160900: contig of 18804 bp in length
          *  160901   161000: gap of unknown length
          *  161001   184481: contig of 23481 bp in length
          *  184482   184581: gap of unknown length
          *  184582   186461: contig of 1880 bp in length
          *  186462   186561: gap of unknown length
          *  186562   191476: contig of 4915 bp in length.
FEATURES             Location/Qualifiers
     source          1..191476
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-438H24"
BASE COUNT    54096 a   39810 c   38404 g   58574 t     502 others
ORIGIN


alignment scores:
            Quality:    7.00              Length:      7
```

```
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945b-2_copy_165_173 x AC026108  ..

 Align seq 1/1  to: AC026108  from: 1  to: 191476

        3 ArgLeuProProMetGlyLys 9
          |||||||||||||||||||||
   177559 AGACTCCCCCCATGGGTAAA 177579

seq name: gb:Mgi:AC064874

Seq documentation block:
LOCUS       AC064874   192487 bp    DNA             HTG       07-DEC-2000
DEFINITION  Homo Sapiens chromosome 2 clone RP11-683I6, WORKING DRAFT
            SEQUENCE, 21 unordered pieces.
ACCESSION   AC064874
VERSION     AC064874.1  GI:26327828
KEYWORDS    HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 192487)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 192487)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-APR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT
            ---------- -- Genome Center ------ - -------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
              ---------   Project Information ------ ----
            Center project name: H_NH0680016
              -- ------   Summary Statistics ------ ----
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 183492 bases at least Q40
            Consensus quality: 186530 bases at least Q30
            Consensus quality: 187943 bases at least Q20
            Insert size: 188000; agarose fp
            Insert size: 190487; sum of contigs
            quality coverage: 4.25 in Q20 bases; agarose fp
            quality coverage: 4.22 in Q20 bases; sum of contigs
              ------ -  -------- ------ ------ - - -
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 21 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *       1     1241: contig of 1241 bp in length
          *    1242     1341: gap of unknown length
          *    1342     2584: contig of 1243 bp in length
          *    2585     2684: gap of unknown length
          *    2685     4809: contig of 2125 bp in length
          *    4810     4909: gap of unknown length
          *    4910     6358: contig of 1449 bp in length
          *    6359     6458: gap of unknown length
          *    6459     9432: contig of 2974 bp in length
```

```
            *      9433     9532: gap of unknown length
            *      9533    12436: contig of 2904 bp in length
            *     12437    12536: gap of unknown length
            *     12537    16233: contig of 3697 bp in length
            *     16234    16333: gap of unknown length
            *     16334    22430: contig of 6097 bp in length
            *     22431    22530: gap of unknown length
            *     22531    29123: contig of 6593 bp in length
            *     29124    29223: gap of unknown length
            *     29224    35329: contig of 6106 bp in length
            *     35330    35429: gap of unknown length
            *     35430    43959: contig of 8530 bp in length
            *     43960    44059: gap of unknown length
            *     44060    52614: contig of 8555 bp in length
            *     52615    52714: gap of unknown length
            *     52715    59526: contig of 6812 bp in length
            *     59527    59626: gap of unknown length
            *     59627    67373: contig of 7747 bp in length
            *     67374    67473: gap of unknown length
            *     67474    78810: contig of 11337 bp in length
            *     78811    78910: gap of unknown length
            *     78911    91308: contig of 12398 bp in length
            *     91309    91408: gap of unknown length
            *     91409   103033: contig of 11625 bp in length
            *    103034   103133: gap of unknown length
            *    103134   117294: contig of 14161 bp in length
            *    117295   117394: gap of unknown length
            *    117395   132060: contig of 14666 bp in length
            *    132061   132160: gap of unknown length
            *    132161   158307: contig of 26147 bp in length
            *    158308   158407: gap of unknown length
            *    158408   192487: contig of 34080 bp in length.
FEATURES             Location/Qualifiers
     source          1. .192487
                     /organism="Homo sapiens"
                     /db_xref "taxon:9606"
                     /chromosome="2"
                     /clone "RP11-680O16"
     misc_feature    1. .1241
                     /note="assembly_name:Contig5"
     misc_feature    1342. .2584
                     /note="assembly_name:Contig6"
     misc_feature    2685. .4809
                     /note="assembly_name:Contig7"
     misc_feature    4910. .6358
                     /note="assembly_name:Contig8
                     clone_end:SP6
                     vector_side:right"
     misc_feature    6459. .9432
                     /note="assembly_name:Contig9"
     misc_feature    9533. .12436
                     /note="assembly_name:Contig10"
     misc_feature    12537. .16233
                     /note="assembly_name:Contig11"
     misc_feature    16334. .22430
                     /note="assembly_name:Contig12"
     misc_feature    22531. .29123
                     /note="assembly_name:Contig13"
     misc_feature    29224. .35329
                     /note="assembly_name:Contig14"
     misc_feature    35430. .43959
                     /note="assembly_name:Contig15"
     misc_feature    44060. .52614
                     /note="assembly_name:Contig16"
     misc_feature    52715. .59526
                     /note="assembly_name:Contig17"
     misc_feature    59627. .67373
                     /note="assembly_name:Contig18"
     misc_feature    67474. .78810
                     /note="assembly_name:Contig19"
     misc_feature    78911. .91308
                     /note="assembly_name:Contig20"
     misc_feature    91409. .103033
                     /note="assembly_name:Contig21"
     misc_feature    103134. .117294
                     /note="assembly_name:Contig22"
     misc_feature    117395. .132060
                     /note="assembly_name:Contig23"
     misc_feature    132161. .158307
                     /note="assembly_name:Contig24"
     misc_feature    158408. .192487
                     /note="assembly_name:Contig25"
BASE COUNT     51333 a  44422 c  44483 g  50538 t   2011 others
ORIGIN
                                                    .


alignment_scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-9456-2_COPY_165_173 x AC064874    ..

 Align seq 1/1  to: AC064874  from: 1  to: 192487

        3 ArgLeuProProMetGlyLys 9
          ||||||||||||||||||||
   165647 AGACTACCAACCATGGGAAA 165657

seq_name: gb_htg15:AC073157

seq_documentation_block:
LOCUS       AC073157   240068 bp    DNA             HTG       10-NOV-2000
DEFINITION  Homo Sapiens chromosome 2 clone RP11-230A12, WORKING DRAFT
            SEQUENCE, 67 unordered pieces.
ACCESSION   AC073157
VERSION     AC073157.8  GI:11128169
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 240068)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Ogbola,B., Ali-Osman,F.K., Alton,C.,
            Alsbrooks,S.L., Amaratunga,H.C., Are,J.K., Banks,I., Barbaria,J.,
            Benton,T., Bimage,K., Blankenburg,K., Bonner,D., Breyer,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.F., Bumby,T.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,L.E.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,P.,
            Coyle,M.D., Datherone,S.R., David,R., Davila,M.L., Davis,C.,
            Davy Carroll,L., Federich,R.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Duran-Rocha,S., Durbin,K.J., Earnhart,C., Edar,S., Edwards,C.C.,
            Ekhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Haas,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Holmes,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk S.,, Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Jordah,S., Karlsson,A., Kelly,S., Khan,Z., Kinq,L., Korvah,J.,
            Kovar,C., Krutosie,J., Kuroshi,A., Landry,N., Leal,B., Lewis,L.R.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,F., Martin,E., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., Mcleod,M.P., Meade,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,I., Mohabbat,K.,
            Moram,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Ogoh,M., Okwuonu,G., Oraqunye,N., Oviedo,R., Pace,A., Parton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojulokan,I., Rolle,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoosheri,N.,
            Sisson,I., Soderqren,E., Sonaike,T., Sparks,A., Stanley,R.,
```

Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,E., Tellrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhen,J., Zorrilla,S., Nelson,D. and Gibbs,R.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 240068)

AUTHORS Worley,K.C.

TITLE Direct Submission

JOURNAL Submitted (09-JUN-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On Nov 9, 2000 this sequence version replaced gi:9665821.

- - Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc_help@bcm.tmc.edu

- - Project Information

Center project name: HBKT

Center clone name: RP11 240A12

- - - - Summary Statistics

Sequencing vector: M13; L08821

Chemistry: Dye primer BodiPy; 13% of reads

Chemistry: Dye terminator Big Dye; 87% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 184472 bases at least Q40

Consensus quality: 208722 bases at least Q30

Consensus quality: 224040 bases at least Q20

Estimated insert size: 210936; sum of contigs estimation

Quality coverage: 0x in Q20 bases; agarose fp estimation

Quality coverage: 8x in Q20 bases; sum of contigs estimation

- -

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/projects/human_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 67 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*	1	23412: contig of 23412 bp in length
*	23413	23412: gap of unknown length
*	23413	45806: contig of 12394 bp in length
*	45807	45906: gap of unknown length
*	45907	46740: contig of 10834 bp in length
*	46741	46840: gap of unknown length
*	46841	55235: contig of 8395 bp in length
*	55236	55335: gap of unknown length
*	55336	63956: contig of 8621 bp in length
*	63957	64056: gap of unknown length
*	64057	71619: contig of 7563 bp in length
*	71620	71719: gap of unknown length
*	71720	82817: contig of 11098 bp in length
*	82818	82917: gap of unknown length
*	82918	90034: contig of 7117 bp in length
*	90035	90134: gap of unknown length
*	90135	97244: contig of 7110 bp in length
*	97245	97344: gap of unknown length
*	97345	105538: contig of 8194 bp in length
*	105539	105638: gap of unknown length
*	105639	111082: contig of 5444 bp in length
*	111083	111182: gap of unknown length
*	111183	117706: contig of 6524 bp in length
*	117707	117806: gap of unknown length
*	117807	125686: contig of 7880 bp in length
*	125687	125786: gap of unknown length
*	125787	131085: contig of 5299 bp in length
*	131086	131185: gap of unknown length
*	131186	135605: contig of 4420 bp in length
*	135606	135705: gap of unknown length
*	135706	139861: contig of 4156 bp in length
*	139862	139961: gap of unknown length
*	139962	142976: contig of 3015 bp in length
*	142977	143076: gap of unknown length
*	143077	145413: contig of 2337 bp in length
*	145414	145513: gap of unknown length
*	145514	147814: contig of 2301 bp in length
*	147815	147914: gap of unknown length
*	147915	150628: contig of 2714 bp in length
*	150629	150728: gap of unknown length
*	150729	153720: contig of 2992 bp in length
*	153721	153820: gap of unknown length
*	153821	158444: contig of 4624 bp in length
*	158445	158444: gap of unknown length
*	158445	159885: contig of 1441 bp in length
*	159886	159985: gap of unknown length
*	159986	161886: contig of 1901 bp in length
*	161887	161986: gap of unknown length
*	161987	163655: contig of 1669 bp in length
*	163656	163755: gap of unknown length
*	163756	167041: contig of 3286 bp in length
*	167042	167141: gap of unknown length
*	167142	169640: contig of 2499 bp in length
*	169641	169740: gap of unknown length
*	169741	172073: contig of 2333 bp in length
*	172074	172173: gap of unknown length
*	172174	174709: contig of 2536 bp in length
*	174710	174809: gap of unknown length
*	174810	176490: contig of 1681 bp in length
*	176491	176590: gap of unknown length
*	176591	177978: contig of 1388 bp in length
*	177979	178078: gap of unknown length
*	178079	180355: contig of 2277 bp in length
*	180456	180455: gap of unknown length
*	[illegible]	182202: contig of 2647 bp in length
*	182103	182202: gap of unknown length
*	182203	186474: contig of 4272 bp in length
*	186375	186474: gap of unknown length
*	186475	188194: contig of 1720 bp in length
*	188195	188294: gap of unknown length
*	188295	190146: contig of 1852 bp in length
*	190147	190246: gap of unknown length
*	190247	192650: contig of 2404 bp in length
*	192651	192750: gap of unknown length
*	192751	194593: contig of 1843 bp in length
*	194594	194693: gap of unknown length
*	194694	196620: contig of 1927 bp in length
*	196621	196720: gap of unknown length
*	196721	198457: contig of 1737 bp in length
*	198458	198557: gap of unknown length
*	198558	200273: contig of 1716 bp in length
*	200274	200373: gap of unknown length
*	200374	201841: contig of 1468 bp in length
*	201842	201941: gap of unknown length
*	201942	203462: contig of 1521 bp in length
*	203463	203562: gap of unknown length
*	203563	205360: contig of 1798 bp in length
*	205361	205460: gap of unknown length
*	205461	207083: contig of 1623 bp in length
*	207084	207183: gap of unknown length
*	207184	208918: contig of 1735 bp in length
*	208919	209018: gap of unknown length
*	209019	210991: contig of 1973 bp in length
*	210992	211091: gap of unknown length
*	211092	212549: contig of 1458 bp in length
*	212550	212649: gap of unknown length
*	212650	214147: contig of 1498 bp in length
*	214148	214247: gap of unknown length
*	214248	215972: contig of 1725 bp in length
*	215973	216072: gap of unknown length

```
            *   216073   217381: contig of 1309 bp in length
            *   217382   217481: gap of unknown length
            *   217482   218713: contig of 1232 bp in length
            *   218714   218813: gap of unknown length
            *   218814   220498: contig of 1685 bp in length
            *   220499   220598: gap of unknown length
            *   220599   221945: contig of 1347 bp in length
            *   221946   222045: gap of unknown length
            *   222046   223113: contig of 1068 bp in length
            *   223114   223213: gap of unknown length
            *   223214   224640: contig of 1427 bp in length

alignment_scores:
            Quality:     7.00              Length:       7
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_165_173 x AC073157    .

 Align seq 1/1  to: AC073157  from: 1  to: 24068

        8 ArgLeuProProMetGlyLys 9
          ||||||||||||||||||||
    95206 AGACTTCCTCCCATGGGTAAA 95226
```

OM nt: US-09-126-945B-2_COPY_165_173 [illegible]

Date: Sep 14 [illegible] PM

At [illegible] Results were produced by the [illegible]
Copyright [illegible]

Command line parameters:
-MODEL frame+_p2n.model -DEV=xlp
-Q [illegible]
-DB N_Geneseq_0601 -QFMT fastap -SUFFIX=olip2n.rng -GAPOP 4.500
-GAPEXT 0.050 -MINMATCH 0.100 -LOOPCL=0.000 -LOOPEXT 0.000
-QGAPOP 4.500 -QGAPEXT=0.050 -XGAPOP 60.000 -XGAPEXT 60.000
-FGAPOP 6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT 60.000
-DELOP 6.000 -DELEXT=7.000 -START 1 -MATRIX oligo
-TRANS=human40.cdi -LIST=45 -DOCALIGN 200 -THR_SCORE quality
-THR_MIN 1 -ALIGN=15 -MODE LOCAL -OUTFMT pto -NORM ext -MINLEN 0
-MAXLEN 2000000000 -USER=US09126945_@CGN1_1_308 -NCPU 6 -ICPU 3
-LONGLOG -NO_XLPXY -WAIT -THREADS 1

Search information block:
Query: US-09-126-945B-2_COPY_165_173
Query length: 9
Database: N_Geneseq_0601:*
Database sequences: 730101
Database length: 313950809
Search time (sec): 680.150000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ...
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF21828	+	[illegible]	167.[illegible]	[illegible]	1087	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF53641	+	9.00	164.11	0.3602	1894	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:AAC83261	+	9.00	164.11	0.3602	1894	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:AAC83266	+	9.00	160.22	0.5926	3317	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV72475		6.00	126.05	47.75	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:AAC48374	+	6.00	[illegible]	[illegible]	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC36547	+	6.00	116.70	157.45	494	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC13877	+	6.00	115.26	[illegible]	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC18399	+	6.00	112.81	[illegible]	750	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC40358	+	6.00	112.80	[illegible]	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC51829	+	6.00	112.55	268.02	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA50562	+	6.00	111.75	[illegible]	1017	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA50560	+	6.00	111.68	296.85	1020	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA50555	+	6.00	111.25	315.21	1079	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA50564	+	6.00	110.88	332.03	1144	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA50559	+	6.00	110.67	341.01	1179	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA50556	+	6.00	[illegible]	[illegible]	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA50546	+	6.00	[illegible]	[illegible]	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA50562	+	6.00	110.15	364.81	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA50551	+	6.00	110.11	366.62	1279	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA50552	+	6.00	109.68	387.18	1360	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA50553	+	6.00	109.68	387.19	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1992.DAT:AAQ29517	+	6.00	109.46	398.54	1405	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ41697	+	6.00	109.46	398.54	1405	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ48251	+	6.00	109.46	398.54	1405	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1995.DAT:AAQ94354	+	6.00	109.46	398.54	1405	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV62389	+	6.00	109.46	398.54	1405	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA50550	+	6.00	109.46	398.54	1405	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA50554	+	6.00	109.28	407.60	1441	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC51798	+	6.00	108.98	426.14	1527	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT48582		6.00	107.21	531.59	1943	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX16719	-	6.00	107.21	531.59	1943	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX37725		6.00	102.94	919.03	3598	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX52264		6.00	102.82	[illegible]	3662	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF52206		6.00	102.82	[illegible]	3662	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF72422	-	6.00	102.82	923.54	3662	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX37713		6.00	102.55	966.31	3807	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX52266		6.00	102.12	1.06e+03	4053	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF52208		6.00	102.12	1.06e+03	4053	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF72424		6.00	102.12	1.06e+03	4053	
/SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:AAC84422	-	6.00	102.12	1.06e+03	4053	
/SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT:AAQ05682	-	6.00	100.59	1.26e+03	5051	
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ32334	-	6.00	100.59	1.26e+03	5051	
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ09915	-	6.00	100.59	1.26e+03	5051	
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV63553	-	6.00	100.59	1.26e+03	5051	

seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF21828

seq_documentation_block:
ID AAF21828 standard; DNA; 1087 BP
XX
AC AAF21828;
XX
DT 27-MAR-2001 (first entry)
XX
DE Human breast and ovarian cancer associated antigen gene SEQ ID 215.
XX
KW Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
KW nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
KW antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
KW Addison's disease; allergy; autoimmune haemolytic anaemia;
KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW cardiovascular disorder; wound healing; neurological disease; ds.
XX
OS Homo sapiens.
XX
PN WO200055173-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05881.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI [illegible], Fulco SM;
XX
DR WPI; 2000-[illegible]/58.
DR P-PSDB; AAB58925.
XX
PT New human breast and ovarian cancer associated gene sequences and the
PT polypeptides encoded by these genes, useful in the prevention,
PT treatment and diagnosis of cancer, immune disorders, cardiovascular
PT disorders and neurological diseases -
XX
PS Claim 1; Page 642-643; 1299pp; English.
XX
CC Sequences AAF21[illegible] - AAF22031 represent DNA sequences encoding human
CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
CC associated with breast and ovarian cancer. Included in the invention are
CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
CC isolation and characterisation of the DNA and protein sequences of the
CC invention. The breast and ovarian cancer associated DNA, protein, agonist
CC or antagonist sequences exhibit cytostatic; immunosuppressive;
CC nootropic; neuroprotective, antiviral, antiallergic; hepatotropic;
CC antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
CC antibacterial, antifungal; antiparasitic and cardiant activity. The
CC polynucleotide and protein sequences are used in the diagnosis of cancer,
CC particularly breast and ovarian cancer. The nucleic acid sequences,
CC proteins, antagonists and agonists may also be used in the diagnosis,
CC prevention and treatment of immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis, cardiovascular disorders such as
CC myocardial ischaemia, wound healing, neurological diseases such as
CC cerebral anoxia and epilepsy; and infectious diseases.
XX
SQ Sequence 1087 BP; 215 A; 347 C; 339 G; 182 T; 4 other;

```
alignment scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2 COPY_165_173 x AAF21828  ..

 Align seq 1/1  to: AAF21828  from: 1  to: 1987

        1 GlnTyrArgLeuProProMetGlyLys 9
          ||||||||||||||||||||||||||||
      369 CAATACCGGCTGCCCCCCATGGGCAAG 395

seq name: /cgn2_6/ptodata/geneseq/geneseqp_NA2000.DAT:AAZ50691

seq documentation block:
ID   AAZ50691 standard; cDNA; 1894 BP.
XX
AC   AAZ50691;
XX
DT   23-MAY-2000  (first entry)
XX
DE   Human prostate derived Ets factor cDNA.
XX
KW   Prostate Derived Ets factor; PDEF; human; chromosome 6p21.3; cancer;
KW   loss of heterozygosity; chromosomal translocation; linkage analysis;
KW   cytostatic; cardiant; immunosuppressive; cerebroprotective; fungicide;
KW   antibacterial; vulnerary; neuroprotective; antiparkinsonian; nootropic;
KW   anabolic; antiinflammatory; anorectic; hybridisation probe; forensic;
KW   tumour marker; diagnosis; treatment; prostate cancer; blood coagulation;
KW   autoimmune disorder; haematopoietic; immune/nervous system; stroke;
KW   neoplasm; microbial infection; tissue regeneration; heart attack;
KW   scarring; food additive; preservative; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             416..1423
FT                   /*tag= a
FT                   /product= "Human prostate derived Ets factor"
FT                   /note= "Expressed in prostate epithelium"
XX
PN   WO200006589-A1.
XX
PD   10-FEB-2000.
XX
PF   02-AUG-1999;   99WO-US17470.
XX
PR   31-JUL-1998;   98US-0126945.
XX
PA   (HUMA ) HUMAN GENOME SCI INC.
PA   (BETH ) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI   Libermann TA, Oettgen JP, Kunsch CA, Endress GA, Rosen CA;
XX
DR   WPI; 2000-195255/17.
DR   P-PSDB; AAY45003.
XX
PT   Novel prostate derived polypeptide, polynucleotide useful in
PT   diagnosis, prevention and treatment of prostate cancer, autoimmune
PT   disorders, microbial infections and also as food additive or
PT   preservative
XX
PS   Claim 4; Fig 1; 132pp; English.
XX
CC   The present cDNA sequence encodes a human prostate derived Ets factor
CC   (PDEF), isolated from human prostate carcinoma cDNA library
CC   (ATCC No.203072). PDEF gene expression is highest in tissues involved in
CC   androgen sensitivity, with enriched expression in prostate epithelium
CC   cells. PDEF gene is mapped to the human chromosome 6p21.3
CC   region that is associated with loss of heterozygosity and chromosomal
CC   translocations in various human cancers. PDEF has cytostatic, cardiant,
CC   immunosuppressive, cerebroprotective, fungicide, antibacterial,
CC   vulnerary, neuroprotective, antiparkinsonian, nootropic, anabolic,
CC   antiinflammatory and anorectic activity. PDEF polynucleotides are useful
CC   in linkage analysis as markers, as hybridisation probes for differential
CC   identification of the tissues or cell types and as polymorphic markers
CC   for forensic purposes. PDEF is useful as prostate-specific tumour marker
CC   for the diagnosis and treatment of prostate cancer. PDEF sequences are
CC   useful for treating autoimmune disorders, haematopoietic, blood
CC   coagulation, immune and nervous system disorders, hyperproliferative
CC   disorders like, neoplasms and microbial infections, heart attacks,
CC   stroke, scarring and for tissue regeneration. They are also useful as
CC   food additives or preservatives.
XX
SQ   Sequence 1894 BP; 468 A; 653 C; 571 G; 302 T; 0 other;


alignment scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2 COPY_165_173 x AAZ50691  ..

 Align seq 1/1  to: AAZ50691  from: 1  to: 1894

        1 GlnTyrArgLeuProProMetGlyLys 9
          ||||||||||||||||||||||||||||
     998 CAATACCGGCTGCCCCCCATGGGCAAG 934

seq name: /cgn2_6/ptodata/geneseq/geneseqp_NA2000.DAT:AAC83261

seq_documentation block:
ID   AAC83261 standard; DNA; 1894 BP.
XX
AC   AAC83261;
XX
DT   19-MAR-2001  (first entry)
XX
DE   Gland specific Ets transcription factor (GSEF) nucleotide sequence.
XX
KW   transcription factor; gland-specific Ets transcription factor; GSEF;
KW   metastatic potential; cancer; tumour; metastasis; breast; prostate;
KW   leukaemia; lymphoma; sarcoma; melanoma; chromosome 6p21.1-6p21.3; ds.
XX
OS   Homo sapiens.
XX
PN   WO200070042-A1.
XX
PD   23-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US13173.
XX
PR   14-MAY-1999;   99US-0134112.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Kaufmann J, Xin H, Harrowe G;
XX
DR   WPI; 2001-011192/02.
DR   P-PSDB; AAI49628.
XX
PT   Detecting metastatic and potential metastatic cancerous cells, useful
PT   for diagnosing, prognosing, grading and staging of cancers by detecting
PT   gland specific Ets transcription factor gene product in a biological
PT   sample from a cell
XX
PS   Disclosure; Page 81-83; 95pp; English.
XX
CC   This invention relates to a method for the detection and determination of
CC   the metastatic potential of a cell. The method comprises detecting a
CC   gland specific Ets transcription factor (GSEF) gene product in a test
```

```
CC   sample. Detection of a GSEF gene product in the test sample in amount
CC   lower than that in a normal cell, is indicative of a cell with high
CC   metastatic potential. The method is useful for determining the metastatic
CC   potential of a cell, for the diagnosis and prognosis of cancer as well as
CC   grading and staging of cancers by detecting GSEF expression in a
CC   biological test sample. The method may also be used to monitor patients
CC   having a predisposition to develop a particular cancer. GSEF polypeptides
CC   are useful for producing antibodies, in cancer diagnosis, prognosis,
CC   grading, staging and management of breast and prostate tumours, and in
CC   detecting polymorphisms in the sequence. GSEF genes and proteins are also
CC   useful in gene therapy. GSEF gene product expression levels can be used
CC   in conjunction with any tissue in which an alteration in GSEF gene
CC   product expression levels is associated with development of a
CC   cancer-associated phenotype. Cancers, which can be monitored include
CC   cancers of the prostate, cervix, lung and colon, melanomas, colorectal
CC   adenocarcinomas, Wilm's tumour, retinoblastoma, sarcomas, myosarcomas,
CC   lung carcinomas, leukaemia, and lymphomas. The GSEF gene is located on
CC   human chromosome 6, specifically at 6p21.1-6p21.3. The present sequence
CC   represents the DNA encoding GSEF.
XX
SQ   Sequence 1894 BP; 368 A; 653 C; 571 G; 302 T; 0 other;


alignment_scores:
            Quality:    9.00              Length:      9
              Ratio:   1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_165_173 x AAC83261    ..

 Align seq 1/1  to: AAC83261  from: 1  to: 1894

        1 GlnTyrArgLeuProProMetGlyLys 9
          |||||||||||||||||||||||||||
      908 CAATACCGGCTGCCCCCCATGGGCAAG 934

seq_comment  (1-1894): [Misc. Sequence Description: DNA] [AC AAC83266]

seq_documentation_block:
ID   AAC83266 standard; DNA; 1894 BP.
XX
AC   AAC83266;
XX
DT   16-MAR-2001  (first entry)
XX
DE   Gland-specific Ets transcription factor (GSEF) cDNA sequence.
XX
KW   Transcription factor; gland-specific Ets transcription factor; GSEF;
KW   metastatic potential; cancer; tumour; metastasis; breast; prostate;
KW   leukaemia; lymphoma; sarcoma; melanoma; chromosome 6p21.1-6p21.3; ss.
XX
OS   Homo sapiens.
XX
PN   WO200070092-A1.
XX
PD   23-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US13173.
XX
PR   14-MAY-1999;   99US-0134112.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Kaufmann J,  Xin H,  Harrowe G;
XX
DR   WPI; 2001-041019/05.
DR   P-PSDB; AAB49633.
XX
PT   Detecting metastatic and potential metastatic cancerous cells, useful
PT   for diagnosing, prognosing, grading and staging of cancers by detecting
PT   gland-specific Ets transcription factor gene product in a biological
PT   sample from a cell -
XX
PS   Disclosure; Fig 2; 95pp; English.
XX
CC   This invention relates to a method for the detection and determination of
CC   the metastatic potential of a cell. The method comprises detecting a
CC   gland specific Ets transcription factor (GSEF) gene product in a test
CC   sample. Detection of a GSEF gene product in the test sample in amount
CC   lower than that in a normal cell, is indicative of a cell with high
CC   metastatic potential. The method is useful for determining the metastatic
CC   potential of a cell, for the diagnosis and prognosis of cancer as well as
CC   grading and staging of cancers by detecting GSEF expression in a
CC   biological test sample. The method may also be used to monitor patients
CC   having a predisposition to develop a particular cancer. GSEF polypeptides
CC   are useful for producing antibodies, in cancer diagnosis, prognosis,
CC   grading, staging and management of breast and prostate tumours, and in
CC   detecting polymorphisms in the sequence. GSEF genes and proteins are also
CC   useful in gene therapy. GSEF gene product expression levels can be used
CC   in conjunction with any tissue in which an alteration in GSEF gene
CC   product expression levels is associated with development of a
CC   cancer associated phenotype. Cancers, which can be monitored include
CC   cancers of the prostate, cervix, lung and colon, melanomas, colorectal
CC   adenocarcinomas, Wilm's tumour, retinoblastoma, sarcomas, myosarcomas,
CC   lung carcinomas, leukaemia, and lymphomas. The GSEF gene is located on
CC   human chromosome 6, specifically at 6p21.1-6p21.3. The present sequence
CC   represents the cDNA encoding GSEF.
XX
SQ   Sequence 3317 BP; 710 A; 1026 C; 970 G; 611 T; 0 other;


alignment_scores:
            Quality:    9.00              Length:      9
              Ratio:   1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_165_173 x AAC83266    ..

 Align seq 1/1  to: AAC83266  from: 1  to: 3317

        1 GlnTyrArgLeuProProMetGlyLys 9
          |||||||||||||||||||||||||||
     2307 CAATACCGGCTGCCCCCCATGGGCAAG 2333

seq_comment  (2307-2333): [Misc. Sequence Description: DNA] [AC AAV72409]

seq_documentation_block:
ID   AAV72409 standard; DNA; 129 BP.
XX
AC   AAV72409;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Human cDNA clone DNA35936 detecting primer #2.
XX
KW   Inflammatory cell infiltration; immune response; T cell proliferation;
KW   anti inflammatory; anti-autoimmune; anti-diabetic; spondyloarthropathy;
KW   T cell-mediated disease; spondyloarthropathy; sclerosis; renal disease;
KW   inflammatory myopathy; hemolytic anemia; thrombocytopenia; thyroiditis;
KW   diabetes mellitus; demyelinating polyneuropathy; Guillain-Barre syndrome;
KW   multiple sclerosis; polyneuropathy; hepatitis; cirrhosis; enteropathy;
KW   sclerosing cholangitis; inflammatory bowel disease; Whipple's disease;
KW   skin disease; dermatitis; psoriasis; asthma; allergic rhinitis; tumor;
KW   food hypersensitivity; urticaria; eosinophilic pneumonia; transplant;
KW   idiopathic pulmonary fibrosis; graft rejection; FR0245; human; primer;
KW   ss.
XX
OS   Homo sapiens.
XX
PN   WO9914211-A2.
XX
PD   25-MAR-1999.
XX
PF   17-SEP-1998;   98WO-US19437.
```

```
XX
PR   17-SEP-1997;   97US-0059119.
PR   18-SEP-1997;   97US-0059264.
PR   28-OCT-1997;   97US-0063550.
PR   12-NOV-1997;   97US-0065186.
PR   21-NOV-1997;   97US-0066364.
PR   24-NOV-1997;   97US-0066770.
PR   04-JUN-1998;   98US-0088026.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Foulq S,  Goddard A,  Gurney AL,  Tumas D,  Wood WI;
XX
DR   WPI; 1999-229499/19.
XX
PT   Composition containing novel polypeptide PRO245, its agonist or
PT   antagonist
XX
PS   Disclosure; Fig 40C; 177pp; English.
XX
CC   This invention describes a novel composition containing (apart from a
CC   carrier or excipient), a novel PRO245 polypeptide (I), its agonist or
CC   antagonist, or their fragments, for modulating: (i) infiltration of
CC   inflammatory cells into tissue; (ii) an immune response; or (iii) T cell
CC   proliferation. The composition increases or decreases any of the effects
CC   (i) (iii). The products of the invention have anti inflammatory,
CC   anti autoimmune and anti diabetic activity. (I), and its (ant)agonists
CC   and their fragments, are used to treat immune related diseases,
CC   particularly T cell-mediated diseases. The diseases treated include
CC   systemic lupus erythematosus, rheumatoid arthritis, juvenile chronic
CC   arthritis, spondyloarthropathies, systemic sclerosis (scleroderma),
CC   idiopathic inflammatory myopathies (dermatomyositis, polymyositis),
CC   Sjogren's syndrome, systemic vasculitis, sarcoidosis, autoimmune
CC   hemolytic anemia (immune pancytopenia, paroxysmal nocturnal
CC   hemoglobinuria), autoimmune thrombocytopenia (idiopathic thrombocytopenic
CC   purpura immune-mediated thrombocytopenia), thyroiditis (Grave's disease,
CC   Hashimoto's thyroiditis, juvenile lymphocytic thyroiditis, atrophic
CC   thyroiditis), diabetes mellitus, immune mediated renal disease
CC   (glomerulonephritis, tubulointerstitial nephritis), multiple sclerosis,
CC   idiopathic demyelinating polyneuropathy, Guillain-Barre syndrome, chronic
CC   inflammatory demyelinating polyneuropathy, infectious hepatitis
CC   (hepatitis A, B, C, D, E and other non hepatotropic viruses), autoimmune
CC   chronic active hepatitis, primary biliary cirrhosis, granulomatous
CC   hepatitis, and sclerosing cholangitis, inflammatory bowel disease
CC   (ulcerative colitis, Crohn's disease), gluten sensitive enteropathy, and
CC   Whipple's disease. Autoimmune or immune mediated skin diseases including
CC   bullous skin diseases, erythema multiforme,contact dermatitis, psoriasis,
CC   asthma, allergic rhinitis, atopic dermatitis, food hypersensitivity,
CC   urticaria, eosinophilic pneumonia, idiopathic pulmonary fibrosis,
CC   hypersensitivity pneumonitis, and transplantation associated diseases
CC   (graft rejection, and graft-versus host disease). (I), its (ant)agonists
CC   or fragment can also be used as an adjuvant in treatment of tumors.
CC   Antibodies against (I) can also be used for diagnosing such diseases.
CC   This sequence represents a primer used to detect a nucleic acid sequence
CC   derived from human cDNA clone DNA35936 which is described in the method
CC   of the invention.
XX
SQ   Sequence 129 BP; 47 A; 35 C; 36 G; 21 T; 9 other;


Alignment scores:
                Quality:     6.00              Length:       6
                  Ratio:    1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-126-945B-2 (1-Y) 165_173 x AAV72409 (1-129)  ..

 Align seq 1/1  to reverse of: AAV72409  from: 1  to: 129

        4 ArgLeuProProMetGly 8
          ||||||||||||||||||
      122 CGCTTACCACCAAPGGGC 105

Seq Name: /cgn2_1/ptodata/1/pctopt/PCTOPT_TEMP_1.seq:NA2000:DAT:AAC38079

Seq document first block:
ID   AAC38079 standard; DNA; 268 BP.
XX
AC   AAC38079;
XX
DT   17-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID No: 19718.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
```

PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144642.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145914.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147048.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151404.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152364.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161859.
PR 26-OCT-1999; 99US-0161860.
PR 26-OCT-1999; 99US-0161861.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2_COPY_165_173 x AAC38079 ..

Align seq 1/1 to: AAC38079 from: 1 to: 268

4 LeuProProMetGlyLys 9
||||:||||||||||||
156 TTGCCTCCAATGGTAAG 173

```
Seq name: /tmp1/n_data/pto_seq/pto_seq_copy/AA_new/AA036547

Seq documentation block:
ID   AA036547 standard; DNA; 494 BP.
XX
AC   AA036547;
XX
DT   17-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 14171.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999; 99US-0121825.
PR   05-MAR-1999; 99US-0123186.
PR   09-MAR-1999; 99US-0123548.
PR   23-MAR-1999; 99US-0125788.
PR   25-MAR-1999; 99US-0126264.
PR   29-MAR-1999; 99US-0126785.
PR   01-APR-1999; 99US-0127462.
PR   06-APR-1999; 99US-0128234.
PR   08-APR-1999; 99US-0128714.
PR   16-APR-1999; 99US-0129845.
PR   19-APR-1999; 99US-0130077.
PR   21-APR-1999; 99US-0130449.
PR   23-APR-1999; 99US-0130510.
PR   23-APR-1999; 99US-0130891.
PR   28-APR-1999; 99US-0131449.
PR   30-APR-1999; 99US-0132048.
PR   30-APR-1999; 99US-0132407.
PR   04-MAY-1999; 99US-0132484.
PR   05-MAY-1999; 99US-0132485.
PR   06-MAY-1999; 99US-0132486.
PR   06-MAY-1999; 99US-0132487.
PR   07-MAY-1999; 99US-0132864.
PR   11-MAY-1999; 99US-0134256.
PR   14-MAY-1999; 99US-0134218.
PR   14-MAY-1999; 99US-0134219.
PR   14-MAY-1999; 99US-0134221.
PR   14-MAY-1999; 99US-0134370.
PR   18-MAY-1999; 99US-0134768.
PR   19-MAY-1999; 99US-0134941.
PR   20-MAY-1999; 99US-0135124.
PR   21-MAY-1999; 99US-0135354.
PR   24-MAY-1999; 99US-0135629.
PR   25-MAY-1999; 99US-0136021.
PR   27-MAY-1999; 99US-0136392.
PR   28-MAY-1999; 99US-0136782.
PR   01-JUN-1999; 99US-0137222.
PR   03-JUN-1999; 99US-0137528.
PR   04-JUN-1999; 99US-0137502.
PR   07-JUN-1999; 99US-0137724.
PR   08-JUN-1999; 99US-0138094.
PR   10-JUN-1999; 99US-0138540.
PR   10-JUN-1999; 99US-0138847.
PR   14-JUN-1999; 99US-0139119.
PR   16-JUN-1999; 99US-0139452.
PR   16-JUN-1999; 99US-0139453.
PR   17-JUN-1999; 99US-0139492.
PR   18-JUN-1999; 99US-0139454.
PR   18-JUN-1999; 99US-0139455.
PR   18-JUN-1999; 99US-0139456.
PR   18-JUN-1999; 99US-0139457.
PR   18-JUN-1999; 99US-0139458.
PR   18-JUN-1999; 99US-0139459.
PR   18-JUN-1999; 99US-0139460.
PR   18-JUN-1999; 99US-0139461.
PR   18-JUN-1999; 99US-0139462.
PR   18-JUN-1999; 99US-0139463.
PR   18-JUN-1999; 99US-0139750.
PR   18-JUN-1999; 99US-0139763.
PR   21-JUN-1999; 99US-0139817.
PR   22-JUN-1999; 99US-0139899.
PR   23-JUN-1999; 99US-0140353.
PR   23-JUN-1999; 99US-0140354.
PR   24-JUN-1999; 99US-0140695.
PR   28-JUN-1999; 99US-0140823.
PR   29-JUN-1999; 99US-0140991.
PR   30-JUN-1999; 99US-0141287.
PR   01-JUL-1999; 99US-0141842.
PR   03-JUL-1999; 99US-0142154.
PR   02-JUL-1999; 99US-0142055.
PR   06-JUL-1999; 99US-0142390.
PR   08-JUL-1999; 99US-0142803.
PR   09-JUL-1999; 99US-0142920.
PR   12-JUL-1999; 99US-0142977.
PR   13-JUL-1999; 99US-0143542.
PR   14-JUL-1999; 99US-0143624.
PR   15-JUL-1999; 99US-0144005.
PR   16-JUL-1999; 99US-0144085.
PR   16-JUL-1999; 99US-0144086.
PR   19-JUL-1999; 99US-0144325.
PR   19-JUL-1999; 99US-0144331.
PR   19-JUL-1999; 99US-0144332.
PR   19-JUL-1999; 99US-0144333.
PR   19-JUL-1999; 99US-0144334.
PR   19-JUL-1999; 99US-0144335.
PR   20-JUL-1999; 99US-0144352.
PR   20-JUL-1999; 99US-0144632.
PR   20-JUL-1999; 99US-0144884.
PR   21-JUL-1999; 99US-0144814.
PR   21-JUL-1999; 99US-0145086.
PR   21-JUL-1999; 99US-0145088.
PR   22-JUL-1999; 99US-0145085.
PR   22-JUL-1999; 99US-0145087.
PR   22-JUL-1999; 99US-0145089.
PR   22-JUL-1999; 99US-0145192.
PR   23-JUL-1999; 99US-0145145.
PR   23-JUL-1999; 99US-0145218.
PR   23-JUL-1999; 99US-0145224.
PR   26-JUL-1999; 99US-0145276.
PR   27-JUL-1999; 99US-0145913.
PR   27-JUL-1999; 99US-0145918.
PR   27-JUL-1999; 99US-0145919.
PR   28-JUL-1999; 99US-0145951.
PR   02-AUG-1999; 99US-0146386.
PR   02-AUG-1999; 99US-0146388.
PR   02-AUG-1999; 99US-0146389.
PR   03-AUG-1999; 99US-0147038.
PR   04-AUG-1999; 99US-0147204.
PR   04-AUG-1999; 99US-0147302.
PR   05-AUG-1999; 99US-0147192.
PR   05-AUG-1999; 99US-0147260.
PR   06-AUG-1999; 99US-0147303.
PR   06-AUG-1999; 99US-0147416.
PR   09-AUG-1999; 99US-0147493.
PR   09-AUG-1999; 99US-0147935.
PR   10-AUG-1999; 99US-0148171.
PR   11-AUG-1999; 99US-0148319.
PR   12-AUG-1999; 99US-0148341.
PR   13-AUG-1999; 99US-0148565.
PR   13-AUG-1999; 99US-0148684.
PR   16-AUG-1999; 99US-0149368.
PR   17-AUG-1999; 99US-0149175.
PR   18-AUG-1999; 99US-0149426.
PR   20-AUG-1999; 99US-0149722.
```

```
PR   20-AUG-1999;   99US-0149724.
PR   20-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151448.
PR   01-SEP-1999;   99US-0151940.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154049.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158242.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

alignment_scores:
            Quality:     6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_165_173 x AAC36547

 Align seq 1/1  to: AAC36547  from: 1  to: 464

       4 LeuProProMetGlyLys 9
         ||||||||||||||||||
     151 TTGCCTCCAATGGGIAAG 168

seq_name: "/tmp1/gcg-cdi-a/geneseq/geneseqp_DNA.fsa" : AAC43877

seq_documentation_block:
ID   AAC43877 standard; DNA; 608 BP.
XX
AC   AAC43877;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 40832.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135453.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
```

PR 18 JUN 1999; 99US 0139459.
PR 18 JUN 1999; 99US 0139460.
PR 18 JUN 1999; 99US 0139461.
PR 18 JUN 1999; 99US 0139462.
PR 18 JUN 1999; 99US 0139463.
PR 18 JUN 1999; 99US 0139750.
PR 18 JUN 1999; 99US 0139764.
PR 21 JUN 1999; 99US 0139817.
PR 22 JUN 1999; 99US 0139899.
PR 23 JUN 1999; 99US 0140354.
PR 23 JUN 1999; 99US 0140354.
PR 24 JUN 1999; 99US 0140695.
PR 28 JUN 1999; 99US 0140824.
PR 29 JUN 1999; 99US 0140991.
PR 30 JUN 1999; 99US-0141287.
PR 01 JUL 1999; 99US 0141842.
PR 01 JUL 1999; 99US 0142154.
PR 02 JUL 1999; 99US 0142055.
PR 06 JUL 1999; 99US 0142390.
PR 08 JUL 1999; 99US 0142804.
PR 09 JUL 1999; 99US 0142920.
PR 12 JUL 1999; 99US 0142977.
PR 13 JUL 1999; 99US 0143542.
PR 14 JUL 1999; 99US-0143624.
PR 15 JUL 1999; 99US 0144005.
PR 16 JUL 1999; 99US-0144085.
PR 16 JUL 1999; 99US-0144086.
PR 19 JUL 1999; 99US 0144325.
PR 19 JUL 1999; 99US-0144331.
PR 19 JUL 1999; 99US-0144332.
PR 19 JUL 1999; 99US-0144333.
PR 19 JUL 1999; 99US 0144334.
PR 19 JUL 1999; 99US 0144335.
PR 20 JUL 1999; 99US 0144352.
PR 20 JUL 1999; 99US 0144632.
PR 20 JUL 1999; 99US 0144884.
PR 21 JUL 1999; 99US-0144814.
PR 21 JUL 1999; 99US 0145086.
PR 21 JUL 1999; 99US-0145088.
PR 22 JUL 1999; 99US 0145085.
PR 22 JUL 1999; 99US 0145087.
PR 22 JUL 1999; 99US-0145089.
PR 22 JUL 1999; 99US-0145192.
PR 23 JUL 1999; 99US 0145145.
PR 23 JUL 1999; 99US 0145218.
PR 23 JUL 1999; 99US 0145224.
PR 26 JUL 1999; 99US-0145276.
PR 27 JUL 1999; 99US-0145913.
PR 27 JUL 1999; 99US-0145918.
PR 27 JUL 1999; 99US 0145919.
PR 28 JUL 1999; 99US-0145951.
PR 02 AUG 1999; 99US-0146386.
PR 02 AUG 1999; 99US-0146388.
PR 02 AUG 1999; 99US 0146389.
PR 03 AUG 1999; 99US 0147038.
PR 04 AUG 1999; 99US 0147284.
PR 04 AUG 1999; 99US 0147302.
PR 05 AUG 1999; 99US 0147192.
PR 05 AUG 1999; 99US 0147260.
PR 06 AUG 1999; 99US 0147304.
PR 06 AUG 1999; 99US 0147416.
PR 09 AUG 1999; 99US 0147494.
PR 09 AUG 1999; 99US 0147945.
PR 10 AUG 1999; 99US 0148171.
PR 11 AUG 1999; 99US 0148319.
PR 12 AUG 1999; 99US-0148341.
PR 13 AUG 1999; 99US-0148565.
PR 13 AUG 1999; 99US-0148684.
PR 16 AUG 1999; 99US-0149368.
PR 17 AUG 1999; 99US 0149175.
PR 18 AUG 1999; 99US 0149426.
PR 20 AUG 1999; 99US 0149722.
PR 20 AUG 1999; 99US 0149723.

PR 20 AUG 1999; 99US-0149929.
PR 23 AUG 1999; 99US 0149902.
PR 23 AUG 1999; 99US 0149930.
PR 25 AUG 1999; 99US-0150566.
PR 26 AUG 1999; 99US-0150884.
PR 27 AUG 1999; 99US-0151065.
PR 27 AUG 1999; 99US-0151066.
PR 27 AUG 1999; 99US 0151080.
PR 30 AUG 1999; 99US 0151303.
PR 31 AUG 1999; 99US-0151438.
PR 01 SEP 1999; 99US 0151930.
PR 07 SEP 1999; 99US-0152363.
PR 10 SEP 1999; 99US-0153070.
PR 13 SEP 1999; 99US-0153758.
PR 15 SEP 1999; 99US-0154018.
PR 16 SEP 1999; 99US-0154039.
PR 20 SEP 1999; 99US-0154779.
PR 22 SEP 1999; 99US-0155139.
PR 23 SEP 1999; 99US-0155486.
PR 24 SEP 1999; 99US-0155659.
PR 28 SEP 1999; 99US-0156458.
PR 29 SEP 1999; 99US 0156596.
PR 04 OCT 1999; 99US-0157117.
PR 05 OCT 1999; 99US-0157754.
PR 06 OCT 1999; 99US 0157865.
PR 07 OCT 1999; 99US-0158029.
PR 08 OCT 1999; 99US-0158232.
PR 12 OCT 1999; 99US 0158369.
PR 13 OCT 1999; 99US-0159293.
PR 13 OCT 1999; 99US-0159294.
PR 13 OCT 1999; 99US-0159295.
PR 14 OCT 1999; 99US 0159329.
PR 14 OCT 1999; 99US-0159330.
PR 14 OCT 1999; 99US-0159331.
PR 14 OCT 1999; 99US-0159637.
PR 14 OCT 1999; 99US-0159638.
PR 18 OCT 1999; 99US-0159584.
PR 21 OCT 1999; 99US-0160741.
PR 21 OCT 1999; 99US-0160767.
PR 21 OCT 1999; 99US-0160768.
PR 21 OCT 1999; 99US 0160770.
PR 21 OCT 1999; 99US-0160814.
PR 21 OCT 1999; 99US-0160815.
PR 22 OCT 1999; 99US 0160980.
PR 22 OCT 1999; 99US-0160981.
PR 22 OCT 1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US 0161406.
PR 26 OCT 1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26 OCT 1999; 99US-0161361.
PR 28 OCT 1999; 99US-0161920.
PR 28 OCT 1999; 99US-0161992.
PR 28 OCT-1999; 99US-0161993.
PR 29 OCT 1999; 99US 0162142.

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2 (1-FT Jan 175) x AAC48877 ..

Align Set 1/2: 1 to AAC48877 from 1 to 608

4 Lenin-Pro-Met-Glylys 9
|||| |||||||||||||
197 LPSPLPGAAISGGIAAG 214

seq_text: /cgn2_6/ptodata/1/pubpa/US09_NEW_PUB/seqdata/09126945b2.DAT:AAC48399

```
seq_documentation_block:
ID   AAC48399 standard; DNA; 750 BP.
XX
AC   AAC48399;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 57344.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   26-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139464.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139764.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140354.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140824.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
```

PR 23 AUG 1999; 99US 0149902.
PR 23 AUG 1999; 99US 0149930.
PR 25 AUG 1999; 99US 0150566.
PR 26 AUG 1999; 99US 0150884.
PR 27 AUG 1999; 99US 0151065.
PR 27 AUG 1999; 99US 0151066.
PR 27 AUG 1999; 99US 0151080.
PR 30 AUG 1999; 99US 0151383.
PR 31 AUG 1999; 99US 0151438.
PR 01 SEP 1999; 99US 0151930.
PR 07 SEP 1999; 99US 0152364.
PR 10 SEP 1999; 99US 0153070.
PR 14 SEP 1999; 99US 0153758.
PR 15 SEP 1999; 99US 0154018.
PR 16 SEP 1999; 99US 0154030.
PR 20 SEP 1999; 99US 0154779.
PR 22 SEP 1999; 99US 0155139.
PR 23 SEP 1999; 99US 0155486.
PR 24 SEP 1999; 99US 0155650.
PR 28 SEP 1999; 99US 0156458.
PR 29 SEP 1999; 99US 0156596.
PR 04 OCT 1999; 99US 0157117.
PR 05 OCT 1999; 99US 0157754.
PR 06 OCT 1999; 99US 0157865.
PR 07 OCT 1999; 99US 0158029.
PR 08 OCT 1999; 99US 0158242.
PR 12 OCT 1999; 99US 0158369.
PR 13 OCT 1999; 99US 0159293.
PR 13 OCT 1999; 99US 0159294.
PR 13 OCT 1999; 99US 0159295.
PR 14 OCT 1999; 99US 0159329.
PR 14 OCT 1999; 99US 0159330.
PR 14 OCT 1999; 99US 0159331.
PR 14 OCT 1999; 99US-0159637.
PR 14 OCT-1999; 99US-0159638.
PR 18 OCT 1999; 99US-0159584.
PR 21 OCT 1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21 OCT-1999; 99US 0160768.
PR 21 OCT 1999; 99US-0160770.
PR 21 OCT 1999; 99US 0160814.
PR 21 OCT 1999; 99US 0160815.
PR 22 OCT 1999; 99US-0160980.
PR 22 OCT 1999; 99US 0160981.
PR 22 OCT 1999; 99US 0160989.
PR 25 OCT 1999; 99US-0161404.
PR 25 OCT 1999; 99US-0161405.
PR 25 OCT 1999; 99US 0161406.
PR 26 OCT 1999; 99US 0161359.
PR 26 OCT 1999; 99US-0161360.
PR 26 OCT 1999; 99US 0161361.
PR 28 OCT 1999; 99US 0161920.
PR 28 OCT 1999; 99US 0161992.
PR 28 OCT 1999; 99US-0161994.
PR 29 OCT 1999; 99US 0162142.

Alignment scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

Alignment block:
US-09-126-945B-2-copy_165_173 x AAC48399 .

Align seq 1/1 to: AAC48399 from: 1 to: 750

4 LeuProProMetGlyLys 9
|||||||||||||||||||
238 TTGCCTCCAATGGGTAAG 255

Seq name: /tmp1/pubdata/pqueue/pqtmp/pq_5AA_...AT_AA_1.rng

Seq documentation block:

ID AAC48399 standard; DNA; 751 BP.
XX
AC AAC48399;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID No: 28018.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144642.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 24-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147945.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.

PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151304.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152864.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157754.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161994.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_165_173 x AAC40358 ..

Align seq 1/1 to: AAC40358 from: 1 to: 751

4 LeuProProMetGlyLys 9
||||||||||||||||||
238 TTGCCCCAATGGGTAAG 255

seq_documentation_block:
ID AAC51829 standard; DNA; 899 BP

XX
AC AAC51829;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 69836.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301449.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128244.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135354.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139454.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139764.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142804.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.

```
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151940.
PR   07-SEP-1999;   99US-0152463.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158242.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159648.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161926.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_165_173 x AAC51829    ..

 Align seq 1/1  to: AAC51829  from: 1  to: 899

       4 LeuProProMetGlyLys 9
         ||||||||||||||||||
     162 TTGCCTCCAATGGGTAAG 179

seq_name:  GIDSI Trial Data:Genseq Issue 30 DNA:*** DAT AAA50562;

seq_documentation_block:
ID   AAA50562 standard; DNA; 1017 BP.
XX
AC   AAA50562;
XX
DT   19-DEC-2000  (first entry)
XX
DE   DNA encoding BHV-1 truncated qlycoprotein D-bovine GnRH tetramer fusion.
XX
KW   BHV-1; qlycoprotein D; tmqD; antiqen; bovine respiratory disease;
KW   qonadotropin releasinq hormone; GnRH; cattle; vaccine;
KW   dual immune response; immunoqen; aqression; fertility;
KW   contraceptive; ds.
XX
OS   Chimeric - Bovine herpesvirus type 1.
OS   Chimeric - Bos taurus.
XX
PN   EP1035133-A2.
XX
PD   13-SEP-2000.
XX
PF   14-FEB-2000; 2000EP-0301164.
XX
PR   17-FEB-1999;   99US-0120454.
XX
PA   (PFIZ ) PFIZER PROD INC.
XX
PI   Campos M, Martinod Sk, Durtschi BA, Yule ID;
XX
DR   WPI; 2000-566247/49.
XX
PT   Novel fusion protein for producinq a dual immune response comprises a
PT   peptide analoqous to an endoqenous peptide which is to be inhibited,
PT   connected to a peptide analoqous to an immunoqen from a pathoqen which
PT   infects a vertebrate.
XX
PS   Example 4; Page 54-54; 99pp; Enqlish.
XX
CC   The present sequence is that of an open readinq frame for a 6XHis
CC   leader, truncated mature qlycoprotein D (tmqD) of bovine
CC   herpesvirus type 1 (BHV-1) and a bovine qonadotropin releasinq
CC   hormone tetramer (4GnRH) encoded by baculovirus vector bac-qD-GnRH.
CC   The tmqD-4GnRH fusion protein is an example of fusion proteins -
CC   the invention comprisinq: a first protein that is analoqous to a
CC   protein endoqenously synthesized within a vertebrate, the activity
CC   of which is to be inhibited within the vertebrate, and which is
CC   incapable by itself of elicitinq an effective immunoinhibitory
CC   response in the vertebrate; and a second protein, which is an
CC   immunoqen from a pathoqen capable of infectinq the vertebrate, and
CC   which causes the vertebrate's immune system to recoqnise the first
CC   protein, producinq a response that inhibits the activity of the
CC   first protein, and also protectinq the vertebrate from infection
CC   by the pathoqen when the vertebrate is vaccinated with the fusion
CC   protein. In the present case, the tmqD-4GnRH fusion protein
CC   elicits a dual immune response that is effective in inhibitinq
CC   sexual characteristics (e.q. aqqressive behaviour in males, and
CC   fertility in males and females, the latter providinq a means of
CC   contraception) in cattle, and also in protectinq aqainst infection
CC   by BHV-1, a causative aqent of bovine respiratory disease.
XX
SQ   Sequence 1017 BP; 181 A; 359 C; 343 G; 134 T; 0 other;


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_165_173 x AAA50562    ..

 Align seq 1/1  to: AAA50562  from: 1  to: 1017

       3 ArqLeuProProMetGly 8
         |||||||||||||||||||
     752 CGCCTCCCCCCATGGGG 769
```

Seq name: US-09-126-945B-2 COPY 165 173 x AAA50560

Seq documentation block:
ID AAA50560 standard; DNA; 1020 BP.
XX
AC AAA50560;
XX
DT 19-DEC-2000 (first entry)
XX
DE DNA encoding BHV-1 truncated mature gD antigen.
XX
KW BHV-1; glycoprotein D; tmgD; antigen; bovine respiratory disease;
KW cattle; vaccine; dual immune response; immunogen; mutant; ds.
XX
OS Bovine herpesvirus type 1.
OS Synthetic.
XX
PN EP1035144-A2.
XX
PD 13-SEP-2000.
XX
PF 14-FEB-2000; 2000EP-0301103.
XX
PR 17-FEB-1999; 99US-0120454.
XX
PA (PFIZ-) PFIZER PROD INC.
XX
PI Campos M, Martinod SR, Durtschi BA, Yule HD;
XX
DR WPI; 2000-566924/53.
DR P-PSDB; AAY96088.
XX
PT Novel fusion protein for producing a dual immune response comprises a
PT peptide analogous to an endogenous peptide which is to be inhibited,
PT connected to a peptide analogous to an immunogen from a pathogen which
PT infects a vertebrate
XX
PS Example 1; Page 52; 94pp; English.
XX
CC The present sequence is that of a nucleotide sequence encoding a
CC truncated mature glycoprotein D (tmgD) of bovine herpesvirus type 1
CC (BHV-1) (see AAY96094). BHV-1 tmgD is an immunogen that was selected
CC as a carrier to combine with cattle gonadotropin releasing hormone
CC (GnRH) in fusion proteins (see AAY96089-91 and AAY96093) for use as
CC vaccines that regulate GnRH activity in cattle while simultaneously
CC providing protection against infection by BHV-1, a causative agent
CC of bovine respiratory disease.
XX
SQ Sequence 1020 BP; 180 A; 361 C; 344 G; 135 T; 0 other;

alignment scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2 COPY 165 173 x AAA50560 ..

Align Seq 1/1 to: AAA50560 from: 1 to: 1020

3 ArgLeuProProMetGly 8
||||||||||||||||||
744 CGCCTGCCGCCGATGGG 764

Seq name: /cgn2_6/ptodata/1/pubpaa/US09_126_945B_2_COPY_165_173_P.pep

Seq documentation block:
ID AAA50555 standard; DNA; 1079 BP.
XX
AC AAA50555;
XX

DT 19-DEC-2000 (first entry)
XX
DE DNA encoding BHV-1 truncated gD antigen.
XX
KW BHV-1; glycoprotein D; tgD; antigen; bovine respiratory disease;
KW cattle; vaccine; dual immune response; immunogen; plasmid pCMV-tgD;
KW mutant; protein; ds.
XX
OS Bovine herpesvirus type 1.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 7..1077
FT /*tag= a
FT sig_peptide 7..60
FT /*tag= b
FT mat_peptide 61..1070
FT /*tag= c
XX
PN EP1035144-A2.
XX
PD 13-SEP-2000.
XX
PF 14-FEB-2000; 2000EP-0301103.
XX
PR 17-FEB-1999; 99US-0120454.
XX
PA (PFIZ-) PFIZER PROD INC.
XX
PI Campos M, Martinod SR, Durtschi BA, Yule HD;
XX
DR WPI; 2000-566924/53.
DR P-PSDB; AAY96092.
XX
PT Novel fusion protein for producing a dual immune response comprises a
PT peptide analogous to an endogenous peptide which is to be inhibited,
PT connected to a peptide analogous to an immunogen from a pathogen which
PT infects a vertebrate
XX
PS Example 5; Fig 11A-C; 94pp; English.
XX
CC The present sequence is that of a nucleotide sequence from plasmid
CC pCMV-tgD. It includes a coding sequence for truncated non-mature
CC glycoprotein D (tgD) of bovine herpesvirus type 1 (BHV-1). BHV-1
CC tgD is an immunogen that was selected as a carrier to combine with
CC cattle gonadotropin releasing hormone (GnRH) (see AAY96084) in fusion
CC proteins in order to regulate GnRH activity in cattle while
CC simultaneously providing protection against infection by BHV-1, a
CC causative agent of bovine respiratory disease.
XX
SQ Sequence 1079 BP; 188 A; 377 C; 367 G; 147 T; 0 other;

alignment scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2 COPY 165 173 x AAA50555 ..

Align Seq 1/1 to: AAA50555 from: 1 to: 1079

3 ArgLeuProProMetGly 8
||||||||||||||||||
805 CGCCTGCCGCCGATGGG 823

Seq name: /cgn2_6/ptodata/1/pubpaa/US09_126_945B_2_COPY_165_173_P.pep

Seq documentation block:
ID AAA50564 standard; DNA; 1144 BP.
XX
AC AAA50564;

XX
DT 19-DEC-2000 (first entry)
XX
DE DNA encoding BHV-1 truncated gD-bovine GnRH tetramer fusion.
XX
KW BHV-1; glycoprotein D; tmgD; antigen; bovine respiratory disease;
KW gonadotropin releasing hormone; GnRH; cattle; vaccine;
KW dual immune response; immunogen; agression; fertility;
KW contraceptive; ds.
XX
OS Chimeric - Bovine herpesvirus type 1.
OS Chimeric - Bos taurus.
XX
FH Key Location/Qualifiers
FT CDS 56..1144
FT /*tag= a
XX
PN EP1035133-A2.
XX
PD 13-SEP-2000.
XX
PF 14-FEB-2000; 2000EP-0301103.
XX
PR 17-FEB-1999; 99US-0120454.
XX
PA (PFIZ) PFIZER PROD INC.
XX
PI Campos M, Martinod SR, Durtschi BA, Yule ID;
XX
DR WPI; 2000-566924/53.
DR P-PSDB; AAY96090.
XX
PT Novel fusion protein for producing a dual immune response comprises a
PT peptide analogous to an endogenous peptide which is to be inhibited,
PT connected to a peptide analogous to an immunogen from a pathogen which
PT infects a vertebrate -
XX
PS Disclosure; Page 55; 93pp; English.
XX
CC The present sequence is that of DNA encoding a tmgD-4GnRH fusion
CC protein comprising a bovine herpesvirus type 1 (BHV-1) truncated
CC mature glycoprotein D (tmgD) antigen joined at the C terminus to a
CC a tetramer of bovine gonadotropin releasing hormone (4GnRH).
CC tmgD-4GnRH (see AAY96090) is an example of fusion proteins comprising:
CC a first protein analogous to a protein endogenously synthesized
CC within a vertebrate, the activity of which is to be inhibited within
CC the vertebrate, and which is incapable by itself of eliciting an
CC effective immunoinhibitory response; and a second protein, which is
CC an immunogen from a pathogen capable of infecting the vertebrate,
CC and which causes the vertebrate's immune system to recognise the
CC first protein, producing a response that inhibits the activity of
CC the first protein, and also protecting the vertebrate from infection
CC by the pathogen when the vertebrate is vaccinated with the fusion
CC protein. In the present case, the tmgD 4GnRH fusion protein
CC elicits a dual immune response that is effective in inhibiting
CC sexual characteristics (e.g. aggressive behaviour in males, and
CC fertility in males and females, the latter providing a means of
CC contraception) in cattle, and also in protecting against infection
CC by BHV 1, a causative agent of bovine respiratory disease.
XX
SQ Sequence 1144 BP; 202 A; 495 C; 385 G; 162 T; 0 other;

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_165_173 x AAA50564 ..

Align seq 1/1 to: AAA50564 from: 1 to: 1144

3 AraLeuProProMetGly 8
| | | | | | |
747 cGCCtGCCCGCCGATGGGG 764

```
FROM of: US-09-126-945B-2_COPY_165_173.* [ ... ] NA *  ...

Date: Sep 14, 2001  4:52 PM

About: Results were produced by the SeqWare ... [ ... ]
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=trans_p2n.model -DEV xlp
 -Q /cgn2_1/ptodata/1/SID_spool_US/09_126_945/usnat_21/us-09-126-945b-2_copy_165_173.*_1267_OFF_query_fasta_1_1267
 -DB Issued_Patents_NA -QFMT fastap -SUFFIX olip2n.rni
 -GAP0P=4.500 -GAPEXT=0.050 -MINMATCH 0.100 -LooPCL 0.000
 -LooPEXT=0.000 -QGAP0P 4.500 -QGAPEXT 0.050  XGAP0P 60.000
 -XGAPEXT 60.000 -FGAP0P=6.000 -FGAPEXT 7.000 -YGAP0P 60.000
 -YGAPEXT=60.000 -DELOP=6.000 -DELEXT=7.000  START 1
 -MATRIX=oligo -TRANS human40.cdi -LIST 45  LooALIGN 200
 -THR_SCORE quality -THR_MIN 1 -ALIGN 15 -MODE LOCAL -OUTFMT pir
 -NORM ext  MINLEN 0  MAXLEN 2000000000
 -USER US09126945_COGNI_1_91 -NCPU 6  LCPU 3  LoNGinn -NO_XLFXY
 -WAIT -THREADS=1

Search Information block:
Query: US-09-126-945B-2_COPY_165_173
Query length: 9
Database: Issued_Patents_NA:*
Database sequences: 424599
Database length: 94655562
Search time (sec): 822.720000

WARN: XGAP0P and YGAP0P must be equal. Assuming YGAP0P XGAP0P 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT=60.000


score_list:
Sequence              Strd Orig    ZScore   EScore Len   ! Documentation  ..
/cgn2_6/pt/data/2/ina/6A_COMB seq US-08-[illegible]-A  +    6.00  111.72  1e1.74    1405  !
/cgn2_6/pt/data/2/ina/6A_COMB seq US-08-082-447-5  +    6.00  111.72  1e1.74    1405  !
/cgn2_6/ptodata/2/ina/6A_COMB seq US-08-831-209-11  -    6.00  109.29  177.67    1943  !
/cgn2_6/ptodata/2/ina/6A_COMB seq US-08-829-553-11  -    6.00  109.29  177.67    1943  !
/cgn2_6/ptodata/2/ina/6B_COMB seq US-08-922-[illegible]A-11  -    6.00  109.29  177.67    1943  !
/cgn2_6/ptodata/2/ina/6B_COMB seq US-08-926-707A-11  -    6.00  109.29  177.67    1943  !
/cgn2_6/ptodata/2/ina/6A_COMB seq US-08-936-706A-11  -    6.00  109.29  177.67    1943  !
/cgn2_6/ptodata/2/ina/6A_COMB seq US-08-248-[illegible]  -    6.00  109.29  177.67    1943  !
/cgn2_6/ptodata/2/ina/6B_COMB seq US-08-[illegible]  -    6.00  109.29  177.67    1943  !
/cgn2_6/ptodata/2/ina/6B_COMB seq US-08-230-[illegible]  +    6.00  122.52  129.94    5051  !
/cgn2_6/ptodata/2/ina/6B_COMB seq US-08-[illegible]  +    6.00  122.52  129.94    5051  !
/cgn2_6/ptodata/2/ina/6A_COMB seq US-08-468-[illegible]  +    6.00  122.52  129.94    5051  !
/cgn2_6/ptodata/2/ina/6A_COMB seq US-09-050-596A-1  -    5.00  122.28   44.00      21  !
/cgn2_6/ptodata/2/ina/6A_COMB seq US-09-050-596A-11  -    5.00  122.28   44.00      21  !
/cgn2_6/ptodata/2/ina/6A_COMB seq US-08-133-[illegible]A-12  -    5.00  119.06   45.79      29  !
/cgn2_6/ptodata/2/ina/6B_COMB seq US-08-[illegible]  +    5.00  119.06   45.79      29  !
/cgn2_6/ptodata/2/ina/6B_COMB seq US-08-451-[illegible]  -    5.00  119.06   45.79      29  !
/cgn2_6/ptodata/2/ina/6A_COMB seq US-09-307-265A-8  -    5.00  119.72   47.25      30  !
/cgn2_6/ptodata/2/ina/6A_COMB seq US-08-[illegible]  -    5.00  116.64   70.09      19  !
/cgn2_6/ptodata/2/ina/6A_COMB seq US-08-[illegible]  -    5.00  116.64   70.09      19  !
/cgn2_6/ptodata/2/ina/6B_COMB seq US-08-476-[illegible]  -    5.00  116.64   70.09      19  !
/cgn2_6/ptodata/2/ina/6A_COMB seq US-08-482-[illegible]  -    5.00  116.64   70.09      19  !
/cgn2_6/ptodata/2/ina/PCTUS_COMB seq PCT-US-[illegible]  -    5.00  116.64   70.09      19  !
/cgn2_6/ptodata/2/ina/backfiles1 seq 5240845 14  +    5.00  116.19   71.24      45  ! Date
/cgn2_6/ptodata/2/ina/6A_COMB seq US-08-167-[illegible]  -    5.00  116.00   77.09      51  !
/cgn2_6/ptodata/2/ina/6A_COMB seq US-08-716-[illegible]  +    5.00  116.00   77.09      51  !
/cgn2_6/ptodata/2/ina/6A_COMB seq US-08-447-172A-27  -    5.00  111.27  139.53      90  !
/cgn2_6/pt/data/2/ina/6A_COMB seq US-08-[illegible]  -    5.00  107.62  241.55     168  !
/cgn2_6/ptodata/2/ina/6B_COMB seq US-08-[illegible]  -    5.00  104.78  214.83     199  !
/cgn2_6/ptodata/2/ina/6B_COMB seq US-09-067-[illegible]  +    5.00  104.78  214.83     199  !
/cgn2_6/ptodata/2/ina/6A_COMB seq US-08-[illegible]  -    5.00  104.78  321.25     199  !
/cgn2_6/ptodata/2/ina/6A_COMB seq US-08-700-[illegible]  -    5.00  104.76  321.28     215  !
/cgn2_6/ptodata/2/ina/6A_COMB seq US-09-050-685A-6  -    5.00  104.16  434.13     280  !
/cgn2_6/pt/data/2/ina/6A_COMB seq US-09-[illegible]  -    5.00  104.16  435.96     281  !
/cgn2_6/pt/data/2/ina/6B_COMB seq US-09-[illegible]  +    5.00  104.16  435.96     281  !
/cgn2_6/pt/data/2/ina/6A_COMB seq US-09-050-[illegible]  -    5.00  104.16  442.41     260  !
/cgn2_6/ptodata/2/ina/6B_COMB seq US-08-[illegible]  -    5.00  103.50  474.4      262  !
/cgn2_6/ptodata/2/ina/6B_COMB seq US-09-[illegible]  +    5.00  103.50  477.07     285  !
/cgn2_6/ptodata/2/ina/6A_COMB seq US-09-[illegible]  +    5.00  103.22  481.69     287  !
/cgn2_6/pt/data/2/ina/6A_COMB seq US-09-[illegible]  +    5.00  103.22  481.69     287  !
/cgn2_6/pt/data/2/ina/6A_COMB seq US-08-594-[illegible]  +    5.00  102.76  415.20     317  !
/cgn2_6/pt/data/2/ina/6A_COMB seq US-09-[illegible]  +    5.00  102.55  424.35     326  !
/cgn2_6/pt/data/2/ina/6A_COMB seq US-08-[illegible]  +    5.00  102.55  426.85     326  !
/cgn2_6/pt/data/2/ina/6B_COMB seq US-09-[illegible]  +    5.00  102.40  435.40     333  !
/cgn2_6/pt/data/2/ina/6A_COMB seq US-09-[illegible]  +    5.00  102.40  435.40     333  !

seq_name "/cgn2_6/pt/data/2/ina/6A_COMB seq US-08-035-558-3

seq_documentation_block:
; Sequence 3, Application US/08035558
; Patent No. 5462734
;  GENERAL INFORMATION:
;    APPLICANT:  Letchworth, Geoffrey J.
;    APPLICANT:  Israel, Barbara A.
;    TITLE OF INVENTION:  HERPESVIRUS VACCINE AND METHOD
;    TITLE OF INVENTION:  OF USING SAME
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Quarles and Brady
;      STREET:  P.O BOX 2113
;      STREET:  FIRST WISCONSIN PLAZA
;      CITY:  MADISON
;      STATE:  WISCONSIN
;      COUNTRY:  U.S.A.
;      ZIP:  53701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 800Kb storage
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh
;      SOFTWARE:  Microsoft Word 4.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/035,558
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/607,794
;      FILING DATE:  02 NOV 1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nicholas J. Seay
;      REGISTRATION NUMBER:  27,486
;      REFERENCE/DOCKET NUMBER:  [illegible]
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (608)251-5000
;      TELEFAX:  (608)251-9166
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1405 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    HYPOTHETICAL:  no
;    ANTI-SENSE:  no
;    ORIGINAL SOURCE:
;      ORGANISM:  bovine herpesvirus-1
;      INDIVIDUAL ISOLATE:  Cooper strain
;    FEATURE:
;      NAME/KEY:  coding sequence for gIV glycoprotein
;      LOCATION:  65 to 1339
;      IDENTIFICATION METHOD:  sequence analysis
;    PUBLICATION INFORMATION:
;      AUTHORS:  Zhou, Yaosheng
;      AUTHORS:  Pittenger, David R.
;      AUTHORS:  Babiuk, Lorne A.
;      AUTHORS:  Lamb, Tim J.
;      TITLE:  Molecular Cloning, Sequencing, and Expression of
;      JOURNAL:  Journal of Virology
;      VOLUME:  64
;      ISSUE:  10
;      PAGES:  5142-5152
;      DATE:  October, 1990
```

```
;       RELEVANT RESIDUES IN SEQ ID NO:  3: FROM 85 TO 1339
US-08-045-558-3


alignment scores:
              Quality:    6.00              Length:       6
                Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_165_173 x US-08-045-558-3    ..

 Align Seq 1/1   to: US-08-045-558-3   from: 1   to: 1405

        3 ArgLeuProProMetGly 8
          ||||||||||||||||||
      884 CGCCTGCCCGCCGATGGGG 901

seq_name: /cgn2_6/ptodata/2/pct/US08682847_5_COMB.seq:US-08-682-847-5

seq documentation block:
; Sequence 5, Application US/08682847
; Patent No. 5858989
;  GENERAL INFORMATION:
;     APPLICANT:  BABIUK, LORNE
;     APPLICANT:  VAN DEN HURK, SYLVIA
;     APPLICANT:  ZAMB, TIM
;     APPLICANT:  FITZPATRICK, DAVID
;     TITLE OF INVENTION:  RECOMBINANT BOVINE HERPESVIRUS TYPE 1
;     TITLE OF INVENTION:  POLYPEPTIDES AND VACCINES
;     NUMBER OF SEQUENCES:  8
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  MORRISON & FOERSTER
;       STREET:  755 PAGE MILL ROAD
;       CITY:  PALO ALTO
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  94304-1018
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/682,847
;       FILING DATE:  12-JUL-1996
;       CLASSIFICATION:  536
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  PARK, FREDDIE K.
;       REGISTRATION NUMBER:  35,636
;       REFERENCE/DOCKET NUMBER:  29419-20005.10
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 813-5600
;       TELEFAX:  (415) 494-0792
;       TELEX:  706141
;   INFORMATION FOR SEQ ID NO:  5:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1405 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  85..1335
US-08-682-847-5


alignment scores:
              Quality:    6.00              Length:       6
                Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_165_173 x US-08-682-847-5    ..

 Align Seq 1/1   to: US-08-682-847-5   from: 1   to: 1405

        3 ArgLeuProProMetGly 8
          ||||||||||||||||||
      884 CGCCTGCCCGCCGATGGGG 901

seq_name: /cgn2_6/ptodata/2/pct/US08631200_11_COMB.seq:US-08-631-200-11

seq documentation block:
; Sequence 11, Application US/08631200
; Patent No. 5846040
;  GENERAL INFORMATION:
;     APPLICANT:  Kleyn, Patrick W.
;     APPLICANT:  Moore, Karen J.
;     TITLE OF INVENTION:  COMPOSITIONS FOR THE TREATMENT AND
;     TITLE OF INVENTION:  DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY
;     NUMBER OF SEQUENCES:  59
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Pennie & Edmonds
;       STREET:  1155 Avenue of the Americas
;       CITY:  New York
;       STATE:  New York
;       COUNTRY:  U.S.A.
;       ZIP:  10036-2711
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/631,200
;       FILING DATE:  12-APR-1996
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Colucci, Laura A.
;       REGISTRATION NUMBER:  30,742
;       REFERENCE/DOCKET NUMBER:  7854-057
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 790-9090
;       TELEFAX:  (212) 869-9741/8864
;       TELEX:  66141 PENNIE
;   INFORMATION FOR SEQ ID NO:  11:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1943 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  DNA (genomic)
US-08-631-200-11


alignment scores:
              Quality:    6.00              Length:       6
                Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_165_173 x US-08-631-200-11/rev    ..

 Align Seq 1/1   to reverse of: US-08-631-200-11   from: 1   to: 1943

        3 ArgLeuProProMetGly 8
          ||||||||||||||||||
      952 AGGCTGCCCGCCATGGGG 935

seq_name: /cgn2_6/ptodata/2/pct/US08829554_11_COMB.seq:US-08-829-554-11

seq documentation block:
; Sequence 11, Application US/08829554
; Patent No. 5817762
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Kleyn, Patrick W.
;    APPLICANT:  Moore, Karen J.
;    TITLE OF INVENTION:  COMPOSITIONS FOR THE TREATMENT AND
;    TITLE OF INVENTION:  DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/829,553
;      FILING DATE:  28-MAR-1997
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/631,200
;      FILING DATE:  12-APR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Corazzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7853-057
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-9741/8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1943 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-829-553-11


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-9458-2_COPY_165_173 x US-08-829-553-11/rev   ..

 Align seq 1/1  to reverse of US-08-829-553-11  from: 1  to: 1943

        3 ArgLeuProProMetGly 8
          ||||||||||||||||||
      952 AGGCTGCCCCCATGGGG 935

seq_name  "/n2_f/patentin/ins_58_COMB.seq" 1 1943

seq_documentation_block:
; Sequence 11, Application US/08/829,553
; Patent No. 5861289
;  GENERAL INFORMATION:
;    APPLICANT:  Kleyn, Patrick W.
;    APPLICANT:  Moore, Karen J.
;    TITLE OF INVENTION:  COMPOSITIONS FOR THE TREATMENT AND
;    TITLE OF INVENTION:  DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY
;    NUMBER OF SEQUENCES:  82
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/922,267A
;      FILING DATE:  2-SEP-1997
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/829,553
;      FILING DATE:  28 MAR-1997
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/631,200
;      FILING DATE:  12-APR-1996
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Corazzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7853-085
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-9741/8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1943 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-922-267A-11


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_165_173 x US-08-922-267A-11/rev   ..

 Align seq 1/1  to reverse of US-08-922-267A-11  from: 1  to: 1943

        3 ArgLeuProProMetGly 8
          ||||||||||||||||||
      952 AGGCTGCCCCCATGGGG 935

seq_name  "/n2_f/patentin/ins_58_COMB.seq" 1 1943

seq_documentation_block:
; Sequence 11, Application US/08/922,267A
; Patent No. 5871931
;  GENERAL INFORMATION:
;    APPLICANT:  Kleyn, Patrick W.
;    APPLICANT:  Moore, Karen J.
;    TITLE OF INVENTION:  COMPOSITIONS FOR THE TREATMENT AND
;    TITLE OF INVENTION:  DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY
;    NUMBER OF SEQUENCES:  60
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS DOS
```

```
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/936,707A
;       FILING DATE:  24 SEP 1997
;       CLASSIFICATION:  514
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Corruzzi, Laura A.
;       REGISTRATION NUMBER:  30,742
;       REFERENCE/DOCKET NUMBER:  7853 100
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 790 9090
;       TELEFAX:  (212) 869 9741/8864
;       TELEX:  66141 PENNIE
;   INFORMATION FOR SEQ ID NO:  11:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1943 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA
US-08-936-707A-11


Alignment scores:
              Quality:     6.00                Length:       6
                Ratio:    1.000                  Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

Alignment block:
 US-09-126-945B-2_COPY_165_173 x US-08-936-707A-11/rev   ..

 Align Seq 1/1  to reverse of: US-08-936-707A-11  from:  1  to:  1943

        3 ArgLeuProProMetGly 8
          |||||||||||||||||||
      962 AGGCTGCCTCCTATGGGG 945

Seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-936-706A-11

Seq_documentation_block:
; Sequence 11, Application US/08936706A
; Patent No. 5876919
;  GENERAL INFORMATION:
;     APPLICANT:  Kleyn, Patrick W.
;     APPLICANT:  Moore, Karen J.
;     TITLE OF INVENTION:  COMPOSITIONS FOR THE TREATMENT AND
;     TITLE OF INVENTION:  DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY
;     NUMBER OF SEQUENCES:  60
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Pennie & Edmonds LLP
;       STREET:  1155 Avenue of the Americas
;       CITY:  New York
;       STATE:  New York
;       COUNTRY:  U.S.A.
;       ZIP:  10036-2711
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/936,706A
;       FILING DATE:  24 SEP 1997
;       CLASSIFICATION:  530
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Corruzzi, Laura A.
;       REGISTRATION NUMBER:  30,742
;       REFERENCE/DOCKET NUMBER:  7853 099
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 790 9090
;       TELEFAX:  (212) 869 9741/8864
;       TELEX:  66141 PENNIE
;   INFORMATION FOR SEQ ID NO:  11:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1943 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA
US-08-936-706A-11


Alignment scores:
              Quality:     6.00                Length:       6
                Ratio:    1.000                  Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

Alignment block:
 US-09-126-945B-2_COPY_165_173 x US-08-936-706A-11/rev   ..

 Align Seq 1/1  to reverse of: US-08-936-706A-11  from:  1  to:  1943

        3 ArgLeuProProMetGly 8
          |||||||||||||||||||
      962 AGGCTGCCTCCTATGGGG 945

Seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-248-203-11

Seq_documentation_block:
; Sequence 11, Application US/09248203
; Patent No. 6043346
;  GENERAL INFORMATION:
;     APPLICANT:  Kleyn, Patrick W.
;     APPLICANT:  Moore, Karen J.
;     TITLE OF INVENTION:  COMPOSITIONS FOR THE TREATMENT AND
;     TITLE OF INVENTION:  DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY
;     NUMBER OF SEQUENCES:  60
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Pennie & Edmonds LLP
;       STREET:  1155 Avenue of the Americas
;       CITY:  New York
;       STATE:  New York
;       COUNTRY:  U.S.A.
;       ZIP:  10036-2711
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/248,203
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/936,707
;       FILING DATE:  24 SEP 1997
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Corruzzi, Laura A.
;       REGISTRATION NUMBER:  30,742
;       REFERENCE/DOCKET NUMBER:  7853 100
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 790 9090
;       TELEFAX:  (212) 869 9741/8864
;       TELEX:  66141 PENNIE
;   INFORMATION FOR SEQ ID NO:  11:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1943 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA
US-09-248-203-11


Alignment scores:
              Quality:     6.00                Length:       6
```

```
              Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment_block:
 US-09-126-945B-2_COPY_165_173 x US-09-248-204-11/rev    .

 Align seq 1/1  to reverse of: US-09-248-204-11  from: 1  to: 1943

        3 ArgLeuProProMetGly 8
          ||||||||||||||||||||
      952 AGGCTGCCTCCTATGGGG 935

Seq_name: /tmp_fp_data_files_FP_COMB_seq:US-09-406-071-11

Seq_documentation_block:
; Sequence 11, Application US/09406071
; Patent No. 6207386
;  GENERAL INFORMATION:
;    APPLICANT:  Kleyn, Patrick W.
;    APPLICANT:  Moore, Karen J.
;    TITLE OF INVENTION:  COMPOSITIONS FOR THE TREATMENT AND
;    TITLE OF INVENTION:  DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY
;    NUMBER OF SEQUENCES:  60
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/406,071
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/936,707
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Corozzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7058.1CC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-9741/8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1943 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-09-406-071-11


Alignment_scores:
            Quality:    6.00                Length:      6
              Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment_block:
 US-09-126-945B-2_COPY_165_173 x US-09-406-071-11/rev    .

 Align seq 1/1  to reverse of: US-09-406-071-11  from: 1  to: 1943

        3 ArgLeuProProMetGly 8
          ||||||||||||||||||||
      952 AGGCTGCCTCCTATGGGG 935
```

```
Seq_name: /tmp_fp_data_files_FP_COMB_seq:US-08-239-276-2

Seq_documentation_block:
; Sequence 2, Application US/08239276
; Patent No. 5840836
;  GENERAL INFORMATION:
;    APPLICANT:  Rabin, Daniel
;    TITLE OF INVENTION:  PANCREATIC ISLET CELL ANTIGENS
;    TITLE OF INVENTION:  OBTAINED BY MOLECULAR CLONING
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Spencer Kramer Schaefer & Briscoe
;      STREET:  660 White Plains Road
;      CITY:  Tarrytown
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10591-5144
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4 Mb storage
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  System 7.5
;      SOFTWARE:  WordPerfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/239,276
;      FILING DATE:  05-MAY-1994
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/872,646
;      FILING DATE:  08-JUN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/715,181
;      FILING DATE:  14-JUN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/441,703
;      FILING DATE:  04 DEC 1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/312,543
;      FILING DATE:  17 FEB 1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kurt G. Briscoe
;      REGISTRATION NUMBER:  33,141
;      REFERENCE/DOCKET NUMBER:  89 : 271.4 EGB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (914) 332 1700
;      TELEFAX:  (914) 332 1844
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5051 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-239-276-2


Alignment_scores:
            Quality:    6.00                Length:      6
              Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment_block:
 US-09-126-945B-2_COPY_165_173 x US-08-239-276-2/rev    .

 Align seq 1/1  to reverse of: US-08-239-276-2  from: 1  to: 5051

        3 ArgLeuProProMetGly 8
          ||||||||||||||||||||
     1430 AGGCTGCCTCCAATGGGG 1413

Seq_name: /tmp_fp_data_files_FP_COMB_seq:US-08-468-579B-2

Seq_documentation_block:
; Sequence 2, Application US/08468579B
```

```
; Patent No. 5981700
;  GENERAL INFORMATION:
;     APPLICANT:  Rabin, Daniel
;     TITLE OF INVENTION:  PANCREATIC ISLET CELL ANTIGENS
;     TITLE OF INVENTION:  OBTAINED BY MOLECULAR CLONING
;     NUMBER OF SEQUENCES:  19
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Sprung Kramer Schaefer & Briscoe
;       STREET:  660 White Plains Road
;       CITY:  Tarrytown
;       STATE:  New York
;       COUNTRY:  USA
;       ZIP:  10591 5144
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 3.50 inch, 1.4 Mb storage
;       COMPUTER:  Apple Macintosh
;       OPERATING SYSTEM:  System 7.5
;       SOFTWARE:  WordPerfect
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/468,577B
;       FILING DATE:  06 JUN 1995
;       CLASSIFICATION:  530
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/239,276
;       FILING DATE:  05 MAY 1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/872,646
;       FILING DATE:  08 JUN 1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/715,181
;       FILING DATE:  14 JUN 1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/441,703
;       FILING DATE:  04 DEC 1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/312,543
;       FILING DATE:  17 FEB 1989
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Kurt G. Briscoe
;       REGISTRATION NUMBER:  33,141
;       REFERENCE/DOCKET NUMBER:  MDI 251.8 KGB
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (914) 332 1700
;       TELEFAX:  (914) 332 1844
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  5051 nucleotides
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US 08 468 577B 2


Alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment_block:
 US 09 126 945B 2_COPY_165_173 x US 08 468 577B 2/rev

 Align seq 1/1  to reverse of: US 08 468 577B 2  from: 1  to: 5051

        3 ArgLeuProProMetGly 8
          |||||||||||||||||
    1430 AGGCTTCCTCCAATGGGC 1413

seq name: /tmp/ftpdata/Isq7A.Mm.seq      3 bases, 0 A, 0 C, 3 G, 0 T

seq documentation block:
; Sequence 2, Application US/08468577B
; Patent No. 6001804
;  GENERAL INFORMATION:
;     APPLICANT:  Rabin, Daniel
;     TITLE OF INVENTION:  PANCREATIC ISLET CELL ANTIGENS
;     TITLE OF INVENTION:  OBTAINED BY MOLECULAR CLONING
;     NUMBER OF SEQUENCES:  19
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Sprung Kramer Schaefer & Briscoe
;       STREET:  660 White Plains Road
;       CITY:  Tarrytown
;       STATE:  New York
;       COUNTRY:  USA
;       ZIP:  10591 5144
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 3.50 inch, 1.4 Mb storage
;       COMPUTER:  Apple Macintosh
;       OPERATING SYSTEM:  System 7.5
;       SOFTWARE:  WordPerfect
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/468,577B
;       FILING DATE:  06 JUN 1995
;       CLASSIFICATION:  514
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/239,276
;       FILING DATE:  05 MAY 1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/872,646
;       FILING DATE:  08 JUN 1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/715,181
;       FILING DATE:  14 JUN 1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/441,703
;       FILING DATE:  04 DEC 1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/312,543
;       FILING DATE:  17 FEB 1989
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Kurt G. Briscoe
;       REGISTRATION NUMBER:  33,141
;       REFERENCE/DOCKET NUMBER:  MDI 251.8 KGB
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (914) 332 1700
;       TELEFAX:  (914) 332 1844
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  5051 nucleotides
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US 08 468 577B 2


Alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment_block:
 US 09 126 945B 2_COPY_165_173 x US 08 468 577B 2/rev

 Align seq 1/1  to reverse of: US 08 468 577B 2  from: 1  to: 5051

        3 ArgLeuProProMetGly 8
          |||||||||||||||||
    1430 AGGCTTCCTCCAATGGGC 1413

seq name: /tmp/ftpdata/Isq7A.Mm.seq      3 bases, 0 A, 0 C, 3 G, 0 T

; Sequence 1, Application US/09050585A
; Patent No. 6033858
;  GENERAL INFORMATION:
;     APPLICANT:  Bastian, Frank O.
;     TITLE OF INVENTION:  DETECTION OF TRANSMISSIBLE SPONGIFORM
```

```
;    TITLE OF INVENTION:  ENCEPHALOPATHIES
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Braman & Rogalskyj, LLP
;      STREET:  P.O. Box 352
;      CITY:  Canandaigua
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  14424-0352
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/050,585A
;      FILING DATE:  30-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Braman, Susan J
;      REGISTRATION NUMBER:  34,103
;      REFERENCE/DOCKET NUMBER:  018.00020
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  716-393-3002
;      TELEFAX:  716-393-3001
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  21 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-050-585A-1


alignment_scores:
            Quality:    5.00              Length:       5
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_165_173 x US-09-050-585A-1/rev

 Align seq 1/1  to reverse of: US-09-050-585A-1  from: 1  to: 21

        3 ArgLeuProProMet 7
          |||||||||||||||
       16 CGCTTGCCACCIATG 2

seq_name: Sun, 6 Apr 2003, /projects/2_Inq/7A_TMP_seq_US-09-050-585A-1

seq_documentation_block:
; Sequence 11, Application US/09050585A
; Patent No. 6033858
;  GENERAL INFORMATION:
;    APPLICANT:  Bastian, Frank O
;    TITLE OF INVENTION:  DETECTION OF TRANSMISSIBLE SPONGIFORM
;    TITLE OF INVENTION:  ENCEPHALOPATHIES
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Braman & Rogalskyj, LLP
;      STREET:  P.O. Box 352
;      CITY:  Canandaigua
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  14424-0352
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/050,585A
;      FILING DATE:  30-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Braman, Susan J
;      REGISTRATION NUMBER:  34,103
;      REFERENCE/DOCKET NUMBER:  018.00020
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  716-393-3002
;      TELEFAX:  716-393-3001
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  21 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-050-585A-11


alignment_scores:
            Quality:    5.00              Length:       5
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_165_173 x US-09-050-585A-11/rev

 Align seq 1/1  to reverse of: US-09-050-585A-11  from: 1  to: 21

        3 ArgLeuProProMet 7
          |||||||||||||||
       16 CGCTTGCCACCIATG 2

seq_name: Sun, 6 Apr 2003, /projects/2_Inq/7A_TMP_seq_US-09-050-585A-11

seq_documentation_block:
; Sequence 12, Application US/08133979A
; Patent No. 5589582
;  GENERAL INFORMATION:
;    APPLICANT:  Hawley, Robert J.
;    APPLICANT:  Ponath, Paul D.
;    APPLICANT:  Rosa, Margaret D.
;    APPLICANT:  Monroy, Rodney L.
;    APPLICANT:  Schafer, Bernice Z.
;    TITLE OF INVENTION:  Enhancement of Xenograft Tolerance and Porcine Cytokines In
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Carella, Byrne, Bain, Gilfillan, Cecchi,Stewart & Olstein
;      STREET:  6 Becker Farm Road
;      CITY:  Roseland
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch diskette
;      COMPUTER:  IBM PC Compatible
;      OPERATING SYSTEM:  MS DOS 5.0
;      SOFTWARE:  Word Perfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/133,979A
;      FILING DATE:  October 8, 1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Herron, Charles J.
;      REGISTRATION NUMBER:  28,019
;      REFERENCE/DOCKET NUMBER:  61750.79
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201 994-1744
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  29 BASES
;      TYPE:  NUCLEIC ACID
```

```
;       STRANDEDNESS:  SINGLE
;       TOPOLOGY:  LINEAR
;     HYPOTHETICAL:  NO
;     ANTI SENSE:  NO
US 08 133 979A 12


Alignment scores:
              Quality:     5.00              Length:        5
                Ratio:    1.000                Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

Alignment block:
 US 09 126 945B 2 COPY 165 173 x US 08 133 979A 12

 Align Seq 1/1  to: US 08 133 979A 12  from: 1  to: 29

        2 TyrArgGlnProPro 6
          ||||||||||||||||
        7 TACCGGCTCCCACCC 21
```

```
M of: US-09-126-945B-2_COPY_165_173 to: EST:*  -out_format = pts

Date: Sep 14, 2001  7:47 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
  -MODEL=framet_p2n.model -DEV=xlp
  -Q=/cgn2_1/ftp/pub/oliprot/247/US-09-126-945B-2_COPY_165_173.rge -L=/cgn2_1/ftp/pub/EST_plus/est_1.lis
  -DB_EST -QFMT=fastap -SUFFIX=olip2n.rst -GAPOP 4.500
  -GAPEXT 0.050 -MINMATCH=0.100 -LoopPCL 0.000  LoopPEXT 0.000
  -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT 60.000
  -FGAPOP 6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT 60.000
  -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX olinu
  -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE quality
  -THR_MIN 1 -ALIGN 15 -MODE LOCAL -OUTFMT pts  NORM ext  MINLEN 0
  -MAXLEN 2000000000 -USER US09126945_3CGN1_1_4645 -NCPU=6
  -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS 1

Search Information block:
Query: US-09-126-945B-2_COPY_165_173
Query length: 9
Database: EST:*
Database sequences: 10228115
Database length: 431459454
Search time (sec): 9087.640000

WARN: YGAPOP and YGAPOP must be equal. Assuming YGAPOP YGAPOP 60.000
WARN: XGAPEXT and YGAPEXP must be equal. Assuming YGAPEXT XGAPEXT 60.000


score_list:
Sequence              Strd Orig    ZScore    EScore Len   ! Documentation ..
qb_est97:BG179568      +     9.00  179.30   2.2e-09  991  ! BG179568 602328182F1 NIH_MGC_91
qb_est24:AI788018      +     7.00  136.56   173.32   488  ! AI788018 oi16c04.y1 Soares mous
qb_est15:A1055648      +     7.00  134.98   212.18   626  ! A1055648 cou00043J05 Cotton Bol
qb_qss4:CNS03LM1       +     7.00  133.70   249.98   766  ! AL249644 Tetraodon nigroviridis
qb_est91:BF699210      -     7.00  132.58   280.81   915  ! BF699210 602126857F1 NIH_MGC_56
qb_qss18:AQ911083      +     7.00  132.27   309.43   960  ! AQ911083 GS31+02512 Trypanosoma
qb_qss5:CNS05KFW       +     7.00  131.41   335.41  1100  ! AL841393 Tetraodon nigroviridis
qb_est74:BF121621            7.00  130.92   357.05  1198  ! BF121621 HWM0126A 21r  ITI* HWM
qb_qss5:D86985         -     6.00  127.29   568.60   193  ! D86985 Human mRNA sequence for
qb_est86:BF370908      +     6.00  126.02   669.75   126  ! BF370908 MK0 PN0097 140600 002
qb_est29:AL589457      -     6.00  124.12   954.29   170  ! AL589457 DKFZp451B2119_r1 451
qb_est32:AV078418      -     6.00  124.01   966.48   173  ! AV078418 AV078418 Mus musculus
qb_est111:Z40978       +     6.00  123.94   974.61   175  ! Z40978 BSC21L8052 normalized int
qb_est43:AV130889      +     6.00  122.87  1.0e-03   207  ! AV130889 AV130889 Mus musculus
qb_est68:BB588154      -     6.00  122.84  1.0e-03   208  ! BB588154 BB588154 RIKEN full le
qb_est47:AV427189      +     6.00  122.63  1.0e-03   215  ! AV427189 AV427189 RIKEN full le
qb_qss37:TA344A04P     -     6.00  122.52  1.0e-03   219  ! AL492245 T. brucei sheared geno
qb_est55:BB068522      +     6.00  122.37  1.1e-03   224  ! BB068522 BB068522 RIKEN full-le
qb_est46:AW487029      -     6.00  122.18  1.1e-03   231  ! AW487029 MK0 SI0019 201509 003
qb_est4:AA228937       +     6.00  121.78  1.2e-03   246  ! AA228937 no59h06.r1 NCI_CGAP_Pr
qb_est34:AV210662      -     6.00  121.75  1.2e-03   247  ! AV210662 AV210662 RIKEN full-le
qb_est23:AI700068      -     6.00  121.65  1.2e-03   251  ! AI700068 wf37a07.x1 NCI_CGAP_Ov
qb_est42:AV061540      +     6.00  121.38  1.2e-03   262  ! AV061540 AV061540 Mus musculus
qb_est12:AA485727            6.00  121.33  1.2e-03   264  ! AA485727 od66d02.s1 NCI_CGAP_Gl
qb_est42:AV084272            6.00  121.33  1.2e-03   264  ! AV084272 AV084272 Mus musculus
qb_est49:AV537106      +     6.00  120.98  1.3e-03   279  ! AV537106 AV537106 Arabidopsis t
qb_est81:BE969548      -     6.00  120.98  1.3e-03   279  ! BE969548 601649769R2 NIH_MGC_74
qb_est8:AA497103       +     6.00  120.94  1.3e-03   281  ! AA497103 ae32d06.r1 Gessler Wil
qb_est32:AV087352      -     6.00  120.89  1.3e-03   283  ! AV087352 AV087352 Mus musculus
qb_est54:BB004242      -     6.00  120.89  1.3e-03   283  ! BB004242 BB004242 RIKEN full-le
qb_est37:AV441124      +     6.00  120.83  1.3e-03   286  ! AV441124 AV441124 RIKEN full le
qb_est59:BB211629      +     6.00  120.78  1.3e-03   288  ! BB211629 BB211629 RIKEN full-le
qb_est12:AA836646      -     6.00  120.76  1.3e-03   289  ! AA836646 od66d02.s1 NCI_CGAP_Gl
qb_est45:AW275052            6.00  120.69  1.3e-03   292  ! AW275052 xm64g38.x1 NCI_CGAP_GC
qb_est104:D81588       +     6.00  120.67  1.3e-03   293  ! D81588 HUM175EU5B Human fetal b
qb_est12:AA861459      -     6.00  120.65  1.3e-03   294  ! AA861459 ak21b01.s1 Soares_test
qb_est95:BG006790      +     6.00  120.65  1.3e-03   294  ! BG006790 RC4-CN0227-271100-011
qb_est108:T03386       +     6.00  120.65  1.3e-03   294  ! T03386 IB1457 Infant brain, Ba
qb_est34:AV227714      +     6.00  120.63  1.3e-03   295  ! AV227714 AV227714 RIKEN full le
qb_qss22:AZ117012      -     6.00  120.63  1.3e-03   295  ! AZ117012 SF [illegible] SP6E
qb_est57:BB142469      -     6.00  120.57  1.3e-03   298  ! BB142469 BB142469 RIKEN full
qb_est32:AV082252      -     6.00  120.48  1.4e-03   302  ! AV082252 AV082252 Mus musculus
qb_est39:AV528056      -     6.00  120.46  1.4e-03   303  ! AV528056 AV528056 Arabidopsis
qb_est111:Z29772       -     6.00  120.42  1.4e-03   305  ! Z29772 A11S2.16 Ara A Arabid
qb_qss37:TA366A04P     -     6.00  120.40  1.4e-03   306  ! AL494041 T. brucei sheared q

seq_name: qb_est97:BG179568

seq_documentation_block:
LOCUS        BG179568        991 bp    mRNA            EST       06-FEB-2001
DEFINITION   602328182F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4429413 5',
             mRNA sequence.
ACCESSION    BG179568
VERSION      BG179568.1  GI:12686271
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 991)
  AUTHORS    [illegible]
  TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL    Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: CHTN/ITP
              cDNA Library Preparation: Life Technologies, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LEAM10181  row: d  column: 22
             High quality sequence stop: 655.
FEATURES             Location/Qualifiers
     source          1..991
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4429413"
                     /clone_lib="NIH_MGC_91"
                     /tissue_type="adenocarcinoma, cecal [inc"
                     /lab_host="DH10B (phage resistant)"
                     /note="organ: prostate, Vector: pCMV-SPORT6, Site_1: NotI;
                     Site_2: SalI, Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.4 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT       196 a    451 c    305 g    149 t
ORIGIN


alignment_scores:
            Quality:     9.00              Length:       9
              Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_165_173 x BG179568  ..

 Align_Seq 1/1  to: BG179568  from: 1  to: 991

        1 GlnGlyArgAlaSerProProMetGlyLys 9
          ||||||||||||||||||||||||||||||||||
      956 CAAIAACGGCCGCGGCCCCATGGGCAAG 422

seq_name: qb_est24:AI788018

seq_documentation_block:
LOCUS        AI788018        488 bp    mRNA            EST       02-JUL-1999
DEFINITION   oi16c04.y1 Soares mouse embryo Mus musculus cDNA clone
             IMAGE:2087718 5', mRNA sequence.
ACCESSION    AI788018
```

```
VERSION     AI788018.1  GI:5335734
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 488)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theisiet,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,K., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Marra M/Wash NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, Mo 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:995402
            Seq primer: custom primer used.
FEATURES                Location/Qualifiers
     source             1. .488
                        /organism "Mus musculus"
                        /strain "C57BL"
                        /db_xref "Taxon:10090"
                        /clone "IMAGE:2087718"
                        /clone_lib "Sugano mouse embryo newa"
                        /dev_stage "embryo, 14 dpc"
                        /lab_host "DH10B"
                        /note "Vector: pME18S-FL4; Site_1: BraIII (CACTGTGCG);
                        Site_2: BraIII (CACCATGTG); 1st strand cDNA was primed
                        with an oligo(dT) primer  [ATGTGGCTTTTTTTTTTTTTTTT];
                        double-stranded cDNA was ligated to a  BraIII adaptor
                        [PGTTGGCCTACTGG], digested and cloned into distinct BraIII
                        sites of the pME18S-FL4 vector (5' site CACTGTGCG, 3' site
                        CACCATGTG). XhoI should be used to isolate the cDNA
                        insert.  Size selection was performed to exclude fragments
                        <1.5kb.  Library constructed by Dr. Sumio Sugano
                        (University of Tokyo Institute of Medical Science).
                        Custom primers  for sequencing: 5' end primer
                        CTTCTGCTCTAAAAGCTGCG and 3' end primer
                        CGACCTGCAGCTCGAGCACA."
BASE COUNT      124 a    107 c     94 g    163 t      1 others
ORIGIN


alignment scores:
            Quality:    7.00              Length:      7
              Ratio:   1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945b-2 (COPY 165 173 x AI788018  ..

 Align seq 1/1  to: AI788018  from: 1  to: 488

        8 ArgLeuProProMetGlyLys 9
          |||||||||||||||||||||
      417 CGTTGGCCTCCTATGGGCAAG 437

Seq name: gb_est15:AI055648

Seq documentation block:
LOCUS       AI055648       626 bp    mRNA            EST       16 JUL 1998
DEFINITION  coan0004.I05 Cotton Boll Abscission Zone cDNA Library Gossypium
            hirsutum cDNA clone coan0004.I05 5', mRNA sequence.
ACCESSION   AI055648
VERSION     AI055648.1  GI:3326762
KEYWORDS    EST.
SOURCE      upland cotton.
  ORGANISM  Gossypium hirsutum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE   1  (bases 1 to 626)
  AUTHORS   Wan,C.H., Yu,Y., Sasinowski,M. and Wing,R.A.
  TITLE     Cotton EST Database: Sequence Analysis of 2000 cDNA Clones from an
            Abscission Zone Library
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: SP030 (AACAGCTATGACCATGATTA)
            High-quality sequence stop: 182.
FEATURES                Location/Qualifiers
     source             1. .626
                        /organism "Gossypium hirsutum"
                        /db_xref "taxon:4635"
                        /clone "coan0004.I05"
                        /clone_lib "Cotton Boll Abscission Zone cDNA Library"
                        /lab_host "XL1 Blue MRF'"
                        /note "Vector: Uni ZAP XR; Site 1: EcoRI; Site 2: XhoI;
                        This is a Uni ZAP XR custom cDNA library made by
                        Stratagene (U.S.A.: 1-800 424 9444); Stratagene cat.
                        #937201."
BASE COUNT      173 a    149 c    133 g    171 t
ORIGIN


alignment scores:
            Quality:    7.00              Length:      7
              Ratio:   1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945b-2 (COPY 165 173 x AI055648  ..

 Align seq 1/1  to: AI055648  from: 1  to: 626

       1 GlnLysArgLeuProProMet 7
         |||||||||||||||||||||
     161 CAGTACCGCTTGCCACCAATG 181

Seq name: gb_gss4:CNS03LMI

Seq documentation block:
LOCUS       CNS03LMI       766 bp    DNA             GSS       17-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC ori end of clone
            036C17 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL249634
VERSION     AL249634.1  GI:2970646
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Pereomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 766)
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fischer,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 766)
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fischer,C.,
```

```
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   4  (bases 1 to 766)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12 APR 2000) to the EMBL/GenBank/DDBJ databases.
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetra.html
FEATURES             Location/Qualifiers
     source          1..766
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="036C17"
                     /clone_lib "G"
                     /note="Genoscope sequence ID : C0B6036AB09SP1 end :
                     PUC-ori"
BASE COUNT       192 a    211 c    190 g    159 t     14 others
ORIGIN


alignment_scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_165_173 x CNS03LM1  ..

 Align seq 1/1  to: CNS03LM1  from: 1  to: 766

       1 GlnTyrArgLeuProProMet 7
         |||||||||||||||||||||
     304 CAGTACCGTCTCCCACCCATG 424

seq_name: qb_est91:BF699210

seq_documentation_block:
LOCUS       BF699210       915 bp    mRNA            EST       22-DEC-2000
DEFINITION  602126857F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4283438 5',
            mRNA sequence.
ACCESSION   BF699210
VERSION     BF699210.1  GI:11984618
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 915)
  AUTHORS   NIH_MGC http://www.ncbi.nlm.nih.gov/
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ARCC
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1115  row: b  column: 15
            High quality sequence stop: 588.
FEATURES             Location/Qualifiers
     source          1..915
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4283438"
                     /clone_lib "NIH_MGC_56"
                     /tissue type "primitive neuroectoderm"
                     /lab host "DH10B (T1 phage resistant)"
                     /note="Organ: brain; Vector: pSPORT1; Site_1:
                     SfiI (ggccgacctcgcgg); Site_2: SfiI (ggccgatgcggcc);
                     Double-stranded cDNA was prepared from cell line RNA.  5'
                     and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5' ATTCTAGAGGCCGAGGCGGCCGACATG d(T)30BN-3'
                     (where B=  A, C, or G and N=  A, C, G, or T).  Average
                     insert size 1.65 Kb (range 0.9-4.0 Kb).  15/15 colonies
                     contained inserts by PCR. This library was enriched for
                     full length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT       238 a    186 c    187 g    304 t
ORIGIN


alignment_scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_165_173 x BF699210/rev  ..

 Align seq 1/1  to: reverse of: BF699210  from: 1  to: 915

       3 ArgLeuProProMetGlyLys 9
         |||||||||||||||||||||
     748 CGCCTGCCCACCGATGGGAAAG 728

seq_name: qb_gss18:AQ911083

seq_documentation_block:
LOCUS       AQ911083       960 bp    DNA             GSS       09-JAN-2001
DEFINITION  GSS1G02517 Trypanosoma cruzi random genomic library Trypanosoma
            cruzi genomic clone G2802, DNA sequence.
ACCESSION   AQ911083
VERSION     AQ911083.2  GI:9476308
KEYWORDS    GSS.
SOURCE      Trypanosoma cruzi.
  ORGANISM  Trypanosoma cruzi
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma; Schizotrypanum.
REFERENCE   1  (bases 1 to 960)
  AUTHORS   Aguero,F., Verdun,R., Frasch,A.C.C. and Sanchez,D.O.
  TITLE     A random sequencing approach for the analysis of the trypanosoma
            cruzi genome: general structure, large gene and repetitive DNA
            families, and gene discovery
  JOURNAL   Genome Res. 10 (12), 1996-2005 (2000)
  MEDLINE   20568489
COMMENT     On Jul 21, 2000 this sequence version replaced gi:6499141.
            Contact: Sanchez D.O.
            [illegible] & Tec. Bioq. I [illegible] Nac. de Gral
            San Martin)
            Av. Gral Paz entre Albarellos y Constituyentes, INTI edificio 24
            CP(1650) San Martin, Prov. de BS AS, Argentina
            Tel: 54 11-4580-7255 ext 409
            Fax: 54 11 4752-9639
            Email: dsanchez@fib.unsam.edu.ar
            Seq primer: T7
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1..960
                     /organism="Trypanosoma cruzi"
                     /strain="CL-Brener"
                     /db_xref="taxon:5693"
                     /clone="G2802"
                     /clone_lib "Trypanosoma cruzi random genomic library"
                     /cell_type "epimastigote"
                     /note="Vector: pBS(-) (Stratagene); T. cruzi DNA was
                     randomly sheared using a nebulizer and the 1 to 2 Kb range
                     was gel purified and cloned into the dephosphorylated
```

```
                    HincII site of the vector"
BASE COUNT       294 a    210 c    238 g    286 t      4 others
ORIGIN


alignment scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 DS 09 126 945B 2 COPY_165_173 x AQ011084

 Align seq 1/1  to: AQ011084  from: 1  to: 960

        4 ArgLeuProProMetGlyLys 9
          |||||||||||||||||||||
      420 CGATTACCACCGATGGGAAAA 440

seq name: qb qss5:CNS05KEW

seq documentation block:
LOCUS        CNS05KEW      1100 bp    DNA              GSS       26 MAY 2000
DEFINITION   Tetraodon nigroviridis genome survey sequence 13 end of clone
             024N01 of library A from Tetraodon nigroviridis, genomic survey
             sequence.
ACCESSION    AL441993
VERSION      AL441993.1  GI:8235151
KEYWORDS     GSS; genome survey sequence.
SOURCE       Tetraodon nigroviridis.
  ORGANISM   Tetraodon nigroviridis
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
             Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
             Tetraodontidae; Tetraodon.
REFERENCE    1  (bases 1 to 1100)
  AUTHORS    Roest Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
             Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
             Weissenbach,J.
  TITLE      Characterization and repeat analysis of the compact genome of the
             freshwater pufferfish Tetraodon nigroviridis
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 1100)
  AUTHORS    Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
             Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
             Saurin,W. and Weissenbach,J.
  TITLE      Human gene number estimate provided by genome wide analysis using
             Tetraodon nigroviridis DNA sequence
  JOURNAL    Unpublished
REFERENCE    3  (bases 1 to 1100)
  AUTHORS    Genoscope.
  TITLE      Direct Submission
  JOURNAL    Submitted (12 APR 2000) to the EMBL/GenBank/DDBJ databases
COMMENT      This sequence is a single read and was generated as part of a large
             scale clone end sequencing project of the Tetraodon nigroviridis
             genome. For more information, please take a look at
             http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1..1100
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="024N01"
                     /clone_lib="A"
                     /note="Genoscope sequence ID : C9AA024C9101AI end : T3"
BASE COUNT       291 a    279 c    281 g    233 t     16 others
ORIGIN


alignment scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000
```

```
alignment block:
 DS 09 126 945B 2 COPY 165 173 x CNS05KEW  ..

 Align seq 1/1  to: CNS05KEW  from: 1  to: 1100

        1 GlnIyrArgLeuProProMet 7
          |||||||||||||||||||||
      428 CAATACGCTTCCCACCCATAG 448

seq name: qb qss5:BE421621

seq documentation block:
LOCUS        BE421621      1188 bp    mRNA             EST       24 OCT 2000
DEFINITION   BWM012cA.04: 1183 HWM Barley Leaf Library Bordeum vulgare cDNA
             clone HWM012cA.04, mRNA sequence.
ACCESSION    BE421621
VERSION      BE421621.1  GI:9419464
KEYWORDS     EST.
SOURCE       barley.
  ORGANISM   Hordeum vulgare
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
             ; Triticeae; Hordeum.
REFERENCE    1  (bases 1 to 1188)
  AUTHORS    Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier
             ,S., Dubcovsky,J., Feuillet,C., Gale,M., Gramer,A., Gustafson,P.,
             Hermann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,
             Lamridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y.,
             Sachenai,N., Qualset,C., Schaeh,W., Sekhara),S., Shatilov,M.,
             Sorrells,M., Warburton,M. and Wenzel,G.
  TITLE      International Triticeae EST Cooperative (ITEC): Production of
             Expressed Sequence tags for Species of the Triticeae
  JOURNAL    Unpublished (2000)
COMMENT      Contact: Wenzel G
             :D Munchen, Lehrstuhl fur Pflanzenbau und Pflanzenzuchtung
             Am Hochanger 2, D-85350 Freising Weihenstephan GERMANY
             Fax: 49 08161 71 5173
             Email: wenzel@mm.pbz.agrar.tu-muenchen.de
             International Triticeae EST Cooperative (ITEC)
             http://www.at [illegible]/genome.
FEATURES             Location/Qualifiers
     source          1..1188
                     /organism="Hordeum vulgare"
                     /cultivar="Barke"
                     /db_xref="taxon:4513"
                     /clone="HWM012cA.04"
                     /clone_lib="ITEC HWM Barley Leaf Library"
                     /tissue_type="leaf"
                     /dev_stage="14 day old"
                     /note="Vector: pBluescriptSK( ); 850 bp average insert
                     size."
BASE COUNT       292 a    327 c    339 g    229 t      1 others
ORIGIN


alignment scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 DS 09 126 945B 2 COPY 165 173 x BE421621/rev  ..

 Align seq 1/1  to reverse of: BE421621  from: 1  to: 1188

        2 TyrArgLeuProProMetGly 8
          |||||||||||||||||||||
     1010 TACGGTTACCCCCCATCGGGC 990

seq name: qb qss5:D86886

seq documentation block:
LOCUS        D86886        103 bp    DNA              GSS       07 FEB 1999
```

```
DEFINITION  Human exon sequence in 1.6Mb segment encompassing Down's syndrome
            region, exon, genomic survey sequence.
ACCESSION   D86885
VERSION     D86885.1  GI:1813399
KEYWORDS    GSS; exon trapping for Down's syndrome region.
SOURCE      Homo sapiens DNA, clone:E19-33.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 103)
  AUTHORS   Ohira,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-1996) to the DDBJ/EMBL/GenBank databases. MEKI
            Ohira, Kazusa DNA Research Institute, Laboratory of Gene Structure
            1; 1532-3 Yanauchino, Kisarazu, Chiba 292, Japan
            (E-mail:ohira@kazusa.or.jp, Tel:+81 438 52 3932,
            Fax:+81 438 52-3931)
REFERENCE   2  (sites)
  AUTHORS   Ohira,M., Seki,N., Nagase,T., Ishikawa,H., Suzuki,F., Nomura,N. and
            Ohki,M.
  TITLE     Gene Identification in a 1.6-Mb Region of the Down Syndrome Region
            on Chromosome 21
  JOURNAL   Unpublished (1996)
REFERENCE   3  (sites)
  AUTHORS   Ohira,M., Seki,N., Nagase,T., Suzuki,F., Nomura,N., Ishikawa,H.,
            Hattori,M., Sakaki,Y., Ikeda,I., Muramatsu,Y., Saito,I., Ishikawa,H.
            and Ohki,M.
  TITLE     Gene identification in the 1.6 -Mb of the down syndrome region on
            chromosome 21
  JOURNAL   Genome Res. (1997) In press
FEATURES             Location/Qualifiers
     source          1..103
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /clone="E19-33"
                     /map="21q22.2"
     exon            1..103
                     /note="trapped exon sequence"
BASE COUNT       29 a      13 c      42 g      19 t
ORIGIN


alignment_scores:
            Quality:     6.00              Length:       6
              Ratio:  1.000                Gaps:         0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_165_173 x D86885/rev   ..

 Align seq 1/1  to reverse of: D86885  from: 1  to: 103

       1 GlnTyrArgLeuProPro 6
         ||||||||||||||||||
      89 CAATACGCAGTGCCTCCT 22

seq_name: qb_est86:BF370908

seq_documentation_block:
LOCUS       BF370908       126 bp    mRNA            EST       24-NOV-2000
DEFINITION  MR0-FN0097-140600-002-H19 FN0097 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF370908
VERSION     BF370908.1  GI:11432933
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 126)
  AUTHORS   Dias Neto,E., Correa Correa,R., Verjovski Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brumstein,A., de Oliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the HAESB/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            http://www.ludwig.org.br/cgi-bin/db/dbquery?db=EST&id=MR0-FN0097-
            140600-002-H19&t4 2000-06-14&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 126.
FEATURES             Location/Qualifiers
     source          1..126
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="FN0097"
                     /dev_stage="Adult"
                     /note="organ: prostate_normal; Vector: puc18; Site_1: SmaI
                     ; Site_2: SmaI; A mini-library was made by cloning
                     products derived from  ORESTES PCR (U.S. Letters Patent
                     application No. 196,716 - Ludwig Institute for Cancer
                     Research) profiles into  the pUC 18 vector. Reverse
                     transcription of tissue mRNA and cDNA amplification were
                     performed under low stringency conditions."
BASE COUNT       14 a      46 c      43 g      23 t
ORIGIN


alignment_scores:
            Quality:     6.00              Length:       6
              Ratio:  1.000                Gaps:         0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_165_173 x BF370908   ..

 Align seq 1/1  to: BF370908  from: 1  to: 126

       1 GlnTyrArgLeuProPro 6
         |||| ||||||||||||
      4 CAGTACGGCTGCGGGGG 21

seq_name: qb_est29:AL589457

seq_documentation_block:
LOCUS       AL589457       170 bp    mRNA            EST       07-MAR-2001
DEFINITION  DKFZp451B2119_r1 451 (synonym: hlcc1) spinal cord Homo sapiens cDNA
            clone DKFZp451B2119 5', mRNA sequence.
ACCESSION   AL589457
VERSION     AL589457.1  GI:13243229
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo
REFERENCE   1  (bases 1 to 170)
  AUTHORS   Bloecker,H., Boecher,M., Brandt,P., Mewes,W., Weil,B. and Wiemann
            ,S.
  TITLE     EST (Bloecker,H., Boecher,M., Brandt,P., Mewes,H.W., Weil B. and
            Wiemann,S.)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: bloecker H
            MIPS
```

```
            Am Klopferspitz 18a D-82152 Martinsried, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            Sequenced by GBF (National Research Centre for Biotechnology Ltd.,
            Braunschweig/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            No 3' sequence available.
            This clone (DKFZp451B2119) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin Charlottenburg, GERMANY, Email: clone@rzpd.de.
FEATURES              Location/qualifiers
     source           1..170
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="DKFZp451B2119"
                      /clone_lib="451 (synonym: hlcel) spinal cord"
                      /tissue_type="human spinal cord"
                      /dev_stage="adult"
                      /lab_host="DH10B"
                      /note="Vector: pSport1; Site 1: NotI; Site 2: SalI"
BASE COUNT        40 a      42 c      68 g      30 t
ORIGIN


alignment scores:
            Quality:     6.00              Length:       6
              Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2_COPY_165_173 x AL589457/rev    ..

 Align seq 1/1  to reverse of: AL589457  from: 1  to: 170

       1 GlnTyrArgLeuProPro 6
         ||||||||||||||||||
      62 CAGTACACGTTGCCACCC 45

Seq name: qb_est32:AV078418

Seq documentation block:
LOCUS        AV078418        173 bp    mRNA             EST        24-JUN-1999
DEFINITION   AV078418 Mus musculus stomach C57BL/6J adult Mus musculus cDNA
             clone 2210404B04, mRNA sequence.
ACCESSION    AV078418
VERSION      AV078418.1  GI:5198246
KEYWORDS     EST.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 173)
  AUTHORS    Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
             Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara
             ,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J.,
             Kikuchi,N., Kojima,Y., Matsuyama,I., Nritsuma,H., Oda,H., Owa,C.,
             Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara
             ,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y., Tominaga,N.,
             Watanabe,S., Yagame,M., Yamamura,I., Yokota,T., Yoshino,M.,
             Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
  TITLE      RIKEN Mouse ESTs
  JOURNAL    Unpublished (1999)
COMMENT      Contact: Chie Owa
             Genome Science Laboratory
             RIKEN
             3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
             Tel: 81 298 36 9145
             Fax: 81 298 36 9098
             Email: genome-res@rtc.riken.go.jp
             Thermostabilization and thermoactivation of thermolabile enzymes by
             trehalose and its application for the synthesis of full length cDNA
             (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
             Transcriptional sequencing: A method for DNA sequencing using RNA
             polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):4455-4460 (1998))
             Please visit our web site (http://genome.rtc.riken.go.jp) for
             further details.
FEATURES              Location/qualifiers
     source           1..173
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="2210404B04"
                      /clone_lib="Mus musculus stomach C57BL/6J adult"
                      /sex="male"
                      /tissue_type="stomach"
                      /dev_stage="adult"
BASE COUNT        49 a      34 c      39 g      51 t
ORIGIN


alignment scores:
            Quality:     6.00              Length:       6
              Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2_COPY_165_173 x AV078418/rev    ..

 Align seq 1/1  to reverse of: AV078418  from: 1  to: 173

       4 LeuProProMetGlyLys 9
         ||||||||||||||||||
      95 CTTCCTCCAATGGGAAAA 78

Seq name: qb_est11:Z40978

Seq documentation block:
LOCUS        Z40978          175 bp    mRNA             EST        09-NOV-1994
DEFINITION   HSC2IB052 normalized infant brain cDNA Homo sapiens cDNA clone
             c-2ib05 3', mRNA sequence.
ACCESSION    Z40978
VERSION      Z40978.1  GI:566723
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 175)
  AUTHORS    Aubray,C., Behar,G., Bois,F., Bonchier,C., da Silva,C., Devignes
             ,M.D., Duprat,S., Boulgatte,R., Jumeau,M.N., Lamy,B., Lorenzo,F.,
             Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
             Sebastiani-Kabaktchis,C. and Fossier,A.
  TITLE      IMAGE: molecular integration of the analysis of the human genome
             and its expression
  JOURNAL    C. R. Acad. Sci. III, Sci. Vie 318 (2): 263-272 (1995)
  MEDLINE    95277534
COMMENT      Contact: Genethon
             Genexpress Genethon
             Genethon centre de recherche sur le genome Humain
             1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
             Tel: 33(0)1-6947-2800
             Fax: 33(0)1-6078-6098
             Email: genexpress@genethon.fr
             Single read. 39 T removed at sequence 5'end
             Genexpress_library_id: C; Genexpress sequence_id: clo 2ib05
             Seq primer: (-21)M13 universal.
FEATURES              Location/qualifiers
     source           1..175
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="c-2ib05"
                      /clone_lib="normalized infant brain cDNA"
                      /sex="Female"
                      /tissue_type="total brain"
                      /dev_stage="3 months old"
```

```
                     /note="organ: brain; Vector: latmid BA; Site_1: HindIII;
                     Site_2: NotI; sex Female; dev_stage 3 months old
                     isolate muscular atrophy patient; tissue_type total brain
                     ; total mRNA was oligo-(dT) primed and directionally
                     cloned 5' -> 3' into the HindIII -> NotI sites of the
                     latmid BA vector. Clone library from B.Soares, Psychiatry
                     Dept. Columbia University, USA. Normalization_method:
                     Bento Soares, P.N.A.S in press"
BASE COUNT        43 a      47 c      41 g      37 t       7 others
ORIGIN


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_165_173 x Z40978    ..

 Align seq 1/1  to: Z40978  from: 1  to: 175

       4 LeuProProMetGlyLys 9
         ||||||||||||||||||
     158 CTCCCACCCATGGGAAAA 175

seq_name: gb_est33:AV130889

seq_documentation_block:
LOCUS         AV130889       207 bp    mRNA             EST       01-JUL-1999
DEFINITION  AV130889 Mus musculus C57BL/6J 11 day embryo Mus musculus cDNA
            clone 2700079013, mRNA sequence.
ACCESSION   AV130889
VERSION     AV130889.1  GI:5417124
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 207)
  AUTHORS   Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
            Akuhira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara
            ,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J.,
            Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H., Owa,C.,
            Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara
            ,Y., Suzuki,H., Suzuki,H., Tatono,M., Tomaru,Y., Tomitoma,N.,
            Watanabe,S., Yagame,M., Yamamura,I., Yokota,I., Yoshino,M.,
            Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Chie Owa
            Genome Science Laboratory
            RIKEN
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-9145
            Fax: 81-298-36-9098
            Email: genome-res@rtc.riken.go.jp
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length cDNA
            (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
            Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1..207
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="2700079013"
                     /clone_lib="Mus musculus C57BL/6J 11 day embryo"
                     /sex="mixed"
                     /dev_stage="11-day embryo"

BASE COUNT        56 a      59 c      48 g      38 t       6 others
ORIGIN


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_165_173 x AV130889    ..

 Align seq 1/1  to: AV130889  from: 1  to: 207

       1 GlnTyrArgLeuProPro 6
         ||||||||||||||||||
      84 CAATATAGGCTCCCCCCC 101

seq_name: gb_est68:BB588154

seq_documentation_block:
LOCUS         BB588154       208 bp    mRNA             EST       30-NOV-2000
DEFINITION  BB588154 RIKEN full-length enriched, 6 days neonate skin Mus
            musculus cDNA clone A030005N24 5', mRNA sequence.
ACCESSION   BB588154
VERSION     BB588154.1  GI:11484698
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 208)
  AUTHORS   Aizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T., Carninci,P.
            , Hanagaki,T., Hayatsu,N., Hiraoka,T., Hirozane,T., Hodoyama,Y.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J., Kojima,Y., Konno
            ,H., Kurisaka,M., Matsuyama,T., Miyazaki,A., Nakamura,M., Nishi,K.,
            Nomura,K., Numazaki,R., Okaraki,Y., Okido,T., Owa,C., Sakai,C.,
            Sakai,K., Sasaki,D., Sato,K., Shibata,K., Shibata,Y., Shimadawa,A.,
            Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Tanaka
            ,T., Toya,T., Watahiki,A., Yamamura,I., Yasunishi,A., Yoshida,K.,
            Yoshiki,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Aizawa,K. et al. 2000)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp
            URL:http://genome.rtc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
            Itoh,M., Kiyosaki,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
            Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1..208
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="A030005N24"
```

/clone_lib "RIKEN full length enriched, 6 days neonate skin"
/tissue_type "skin"
/dev_stage "6 days neonate"
/lab_host "DH10B"
/note "Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGACTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo activated reverse transcriptase and subsequently enriched for full length by cap-trapper. cDNA went through one round of normalization to Rot 10.0 and subtraction to Rot 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATCTTCGACTIAATIAAATIAATCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda ZAP II."
BASE COUNT 68 a 40 c 64 g 36 t
ORIGIN

Alignment scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

Alignment block:
US-09-126-945B-2_COPY_165_173 x BB588154.gcg ..

Align seq 1/1 to reverse of: BB588154 from: 1 to: 208

4 LeuProProMetGlyLys 9
|||||||||||||||||||
164 TTACCTCCCATGGCAAG 147

```
OM of: US-09-126-945B-2_COPY_178_186 to: GenEmbl:*   out_format : pts

Date: Sep 14, 2001  9:38 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL frame+_p2n.model -DEV-xlp
 -Q=[illegible]
 -DB GenEmbl -QFMT fastap -SUFFIX olip2n.rge  GAPOP 4.500
 -GAPEXT 0.050 -MINMATCH 0.100 -LOOPCL 0.000  LOOPEXT 0.000
 -QGAPOP 4.500 -QGAPEXT 0.050 -XGAPOP 60.000  XGAPEXT 60.000
 -FGAPOP 6.000 -FGAPEXT 7.000 -YGAPOP 60.000  YGAPEXT 60.000
 -DELOP 6.000 -DELEXT 7.000 -START 1 -MATRIX oligo
 -TRANS human40.cdi -LIST=45 -DOCALIGN 200 -THR_SCORE quality
 -THR_MIN 1 -ALIGN 15 -MODE LOCAL -OUTFMT pts  NORM ext  MINLEN 0
 -MAXLEN 2000000000 -USER US09126945_@CGN1_1_7594 -NCPU 6
 -ICPU 3 -LONGLOG -NO_XLPXY -WAIT -THREADS 1

Search information block:
Query: US-09-126-945B-2_COPY_178_186
Query length: 9
Database: GenEmbl:*
Database sequences: 1344157
Database length: -856060004
Search time (sec): 6503.700000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000


score_list:
Sequence                Strd orig    ZScore   PScore Len   ! Documentation
gb_pat1:AX048148         +     9.00   175.95   0.1454    1894   ! AX048148 Sequence 1 from Paten
gb_pr4:AF071538          +     9.00   175.95   0.1454    1894   ! AF071538 Homo sapiens Ets tran
gb_pr1:AB031549          +     9.00   175.94   0.1457    1898   ! AB031549 Homo sapiens PSF mRNA
gb_pat1.AX048159         +     9.00   172.55   0.2248    3317   ! AX048159 Sequence 12 from Pate
gb_pr6:AL157372                9.00   149.43   4.76      150660 ! AL157372 Human DNA seque
gb_ba3:PSEPCHRA          -     8.00   155.97   1.89      1753   ! L11657 Pseudomonas aeruginosa
gb_ba1:AE004840          +     8.00   145.13   7.57      10481  ! AE004840 Pseudomonas aerugino
gb_ba1:AF074705          -     8.00   142.95   10.02     15023  ! AF074705 Pseudomonas aerugino
gb_htg23:AP000652        +     8.00   130.12   51.91     124808 ! AP000652 Homo sapiens chromo
gb_htg24:AP001876        -     8.00   127.43   73.33     194701 ! AP001876 Homo sapiens chromo
gb_pr9:HSCALRI3          +     7.00   147.08   5.90      260  ! X56670 Human DNA for calretinin
gb_ro1:AB019436          +     7.00   135.69   25.43     1794   ! AB019436 Mus musculus mRNA for
gb_pr6:AL445224          -     7.00   125.14   98.37     9717   ! AL445224 Human DNA sequence fr
gb_pr1:AB031007          -     7.00   121.73   152.36    17063  ! AB031007 Homo sapiens DNA, MH
gb_pr1:AB000879          -     7.00   116.65   292.14    39436  ! AB000879 Homo sapiens genomic
gb_pr1:AC004184          -     7.00   115.89   322.07    44710  ! AC004184 Homo sapiens clone U
gb_htg9:AC024635               7.00   115.56   336.24    47258  ! AC024635 Homo sapiens clone R
gb_htg5:AC016154         +     7.00   115.09   357.04    51053  ! AC016154 Homo sapiens chromos
gb_htg21:AL359978_3      -     7.00   112.18   518.27    82469  ! Continuation (4 of 4) of AL35
gb_pr7:AP000506          -     7.00   111.01   602.02    100000 ! AP000506 Homo sapiens genomi
gb_htg10:AC025519        -     7.00   110.07   679.86    116939 ! AC025519 Homo sapiens clone
gb_ro1:MMBC188A7         -     7.00   109.86   698.09    120990 ! AF109719 Mus musculus casein
gb_pr2:AC006000          +     7.00   109.70   712.36    124170 ! AC006000 Homo sapiens PAC cl
gb_pr4:AC025434          +     7.00   109.41   739.87    130388 ! AC025434 Homo sapiens chromo
gb_pr8:HS453A3           -     7.00   109.38   742.24    130926 ! Z83745 Human DNA sequence fr
gb_pr1:AB045357          +     7.00   109.33   747.54    132131 ! AB045357 Homo sapiens genomi
gb_htg11:AC026734        -     7.00   109.23   756.84    134249 ! AC026734 Homo sapiens chromo
gb_htg11:AC027267        +     7.00   109.19   760.29    135038 ! AC027267 Homo sapiens chromo
gb_htg6:AC018433         -     7.00   109.06   773.91    138133 ! AC018433 Homo sapiens clone
gb_htg10:AC025403        -     7.00   108.85   794.18    142832 ! AC025403 Homo sapiens chromo
gb_pr2:AC005509          +     7.00   108.83   796.59    143391 ! AC005509 Homo sapiens chromo
gb_htg7:AC021269         +     7.00   108.78   801.85    144609 ! AC021269 Homo sapiens clone
gb_htg17:AC083966        -     7.00   108.72   808.36    146122 ! AC083966 Homo sapiens chromo
gb_htg13:AC053479        -     7.00   108.64   816.07    147919 ! AC053479 Homo sapiens chromo
gb_htg9:AC024655         -     7.00   108.62   818.20    148416 ! AC024655 Homo sapiens chromo
gb_htg9:AC024657         +     7.00   108.62   818.62    148513 ! AC024657 Homo sapiens chromo
gb_pr1:AB000882          -     7.00   108.60   820.53    148960 ! AB000882 Homo sapiens genomi
gb_htg13:AC060762        +     7.00   108.36   845.83    154897 ! AC060762 Homo sapiens clone
gb_htg23:AP001528        -     7.00   108.34   848.03    155414 ! AP001528 Homo sapiens chromo
gb_htg20:AL157946        -     7.00   108.31   851.85    156316 ! AL157946 Homo sapiens chr
gb_htg16:AC078992        -     7.00   108.22   861.56    158612 ! AC078992 Homo sapiens chr
gb_htg7:AC021601               7.00   108.14   870.77    160799 ! AC021601 Homo sapiens clo
gb_htg19:AL139215              7.00   108.12   872.25    161149 ! AL139215 Homo sapiens chr
gb_htg14:AC068424              7.00   108.10   875.33    161883 ! AC068424 Mus musculus chr
gb_htg18:AC087446              7.00   108.00   885.80    164378 ! AC087446 Homo sapiens chr

seq_name: gb_pat1:AX048148

seq_documentation_block:
LOCUS       AX048148     1894 bp    DNA             PAT       15-DEC-2000
DEFINITION  Sequence 1 from Patent WO0070092.
ACCESSION   AX048148
VERSION     AX048148.1  GI:11876967
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1894)
  AUTHORS   Kaufmann,J. and Xin,H.
  TITLE     Expression of ets domain proteins in cancer
  JOURNAL   Patent: WO 0070092-A 1 23 NOV 2000;
            Chiron Corporation (US)
FEATURES             Location/Qualifiers
     source          1..1894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             416..1423
                     /note="Human JKETS."
                     /codon_start=1
                     /protein_id="CAC19002.1"
                     /db_xref="GI:11876968"
                     /translation="MCSASPCLSSVSPSP[illegible]FFAAVAVGLERRLWS
                     PSPPAIPEQGLSAFYLSYFDMLYPEDSSWAAKAPGASSREEPPEEPEQCPVIDSQAPA
                     GSLDLVPGGLTLEEHSLEQVQSMVVGEVLKDIETACKLLNITADPMDWSPSNVQKWLL
                     WTEHQYRLPPMGKAFQELAGKELCAMSEEQFRQRSPLGGDVLHAHLDIWKSAAWMKER
                     TSPGAIHYCASTSEESWTDSEVDSSCSGQPIHLWQFLKELLLKPHSYGRFIRWLNKEK
                     GIFKIEDSAQVARLWGIRKNRPAMNYDKLSRSIRQYYKKGIIRKPDISQRLVYQFVHP
                     I"
BASE COUNT      368 a    653 c    571 g    302 t
ORIGIN


alignment scores:
           Quality:     9.00              Length:        9
             Ratio:    1.000                Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US 09 126 945B 2_COPY_178_186 x AX048148  ..

 Align seq 1/1  to: AX048148  from: 1  to: 1894

      1 LeuAlaGlyLysGluLeuCysAlaMet 9
        |||||||||||||||||||||||||||
    947 CTGGCGGGCAAGGAGCTGTGCGCCATG 973

seq_name: gb_pr4:AF071538

seq_documentation_block:
LOCUS       AF071538     1894 bp    mRNA            PRI       10-JAN-2000
DEFINITION  Homo sapiens Ets transcription factor PDEF (PDEF) mRNA, complete
            cds.
ACCESSION   AF071538
VERSION     AF071538.1  GI:4007417
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1894)
  AUTHORS   Oettgen,P., Finger,E., Sun,Z., Akbarali,Y., Thamrongsak,U.,
```

```
            Boltax,J., Stali,F., Fehe,A., Weiss,A., Brown,J., Quinn,G., Kay,E.,
            Endress,G., Kunsch,C. and Libermann,T.A.
  TITLE     PDEF, a novel prostate epithelium specific ets transcription
            factor, interacts with the androgen receptor and activates
            prostate specific antigen gene expression
  JOURNAL   J. Biol. Chem. 275 (2), 1216-1225 (2000)
  MEDLINE   20092894
REFERENCE   2  (bases 1 to 1894)
  AUTHORS   Libermann,T.A., Oettgen,P., Kunsch,C. and Finger,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-1999) Medicine, Beth Israel Deaconess Medical
            Center, 330 Brookline Ave., Boston, MA 02215, USA
FEATURES             Location/Qualifiers
     source          1..1894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6p21.3"
                     /tissue_type="prostate"
     gene            1..1894
                     /gene="PDEF"
     CDS             416..1423
                     /gene="PDEF"
                     /note="prostate epithelium specific"
                     /codon_start=1
                     /product="Ets transcription factor PDEF"
                     /protein_id="AAC99296.1"
                     /db_xref="GI:4507416"
                     /translation="MESASSGLSSVCPSHELLPPEFVSRTGLEKAAAGAVGLEPMLWS
                     PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAPGASSKEEPEPEEQCPVLDSGAPA
                     GSLDLVPGGLLELEHSLEQVQSMVVCEVLKETETACKLENTIADFMEWGFSNVEWLL
                     WFEHQYRLLPEMGKAFQELAGKELCAMSEEQERQPSRLGGDVLHAHLLDWFSAAWMEEM
                     TSPGATHYCASTSEESWTDSEVDSSCSGQFCHLWQELEELLEPDHSYGRPLPWLNKEK
                     GIFKIESVAQVAKLWGLRENKPAMNYDKLSRSLRQYYEKGLIRKPDTSQKLVYQFVHP
                     I"
BASE COUNT       368 a    653 c    571 g    302 t
ORIGIN


alignment scores:
            Quality:    9.00            Length:      9
              Ratio:   1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_178_186 x AF071538  ..

 Align seq 1/1  to: AF071538  from: 1  to: 1894

        1 LeuAlaGlyLysGluLeuCysAlaMet 9
          |||||||||||||||||||||||||||
      947 CTGGCGGGCAAGGAGCTGTGCGCCATG 973

seq_name: gb_pri:AB031549

seq_documentation_block:
LOCUS        AB031549      1898 bp    mRNA            PRI       20-JAN-2000
DEFINITION   Homo sapiens PSE mRNA for prostate ets, complete cds.
ACCESSION    AB031549
VERSION      AB031549.1  GI:6721497
KEYWORDS     prostate ets.
SOURCE       Homo sapiens adult male prostate epithelium prostate cancer
             cell line:PC-3 cDNA to mRNA.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (sites)
  AUTHORS    Yamada,N., Tamai,Y., Miyamoto,H. and Nozaki,M.
  TITLE      Cloning and expression of the mouse Pse gene encoding a novel Ets
             family member
  JOURNAL    Gene 241 (2), 267-274 (2000)
  MEDLINE    20137504
REFERENCE    2  (bases 1 to 1898)
  AUTHORS    Nozaki,M.
  TITLE      Direct Submission
  JOURNAL    Submitted (27-AUG-1999) to the EMBL/GenBank/DDBJ databases. Masami
             Nozaki, Osaka University, Institute for Microbial Diseases; 3-1,
             Yamadaoka, Suita, Osaka 565-0871, Japan
             (E-mail:mnozaki@biken.osaka-u.ac.jp, Tel:81-6-879-8338,
             Fax:81-6-879-8339)
FEATURES             Location/Qualifiers
     source          1..1898
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="PC-3"
                     /cell_type="prostate cancer"
                     /chromosome="6"
                     /dev_stage="adult"
                     /map="6p21.2-21.3"
                     /sex="male"
                     /tissue_type="prostate epithelium"
                     /note="vector:pBC18"
     gene            406..1413
                     /gene="PSE"
     CDS             406..1413
                     /gene="PSE"
                     /codon_start=1
                     /product="prostate ets"
                     /protein_id="BAA89543.1"
                     /db_xref="GI:6721498"
                     /translation="MESASSGLSSVCPSHELLPPEFVSRTGLEKAAAGAVGLEPMLWS
                     PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAPGASSKEEPEPEEQCPVLDSGAPA
                     GSLDLVPGGLLELEHSLEQVQSMVVCEVLKETETACKLENTIADFMEWGFSNVEWLL
                     WFEHQYRLLPEMGKAFQELAGKELCAMSEEQERQPSRLGGDVLHAHLLDWFSAAWMEEM
                     TSPGATHYCASTSEESWTDSEVDSSCSGQFCHLWQELEELLEPDHSYGRPLPWLNKEK
                     GIFKIESVAQVAKLWGLRENKPAMNYDKLSRSLRQYYEKGLIRKPDTSQKLVYQFVHP
                     I"
BASE COUNT       380 a    651 c    567 g    300 t
ORIGIN


alignment scores:
            Quality:    9.00            Length:      9
              Ratio:   1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_178_186 x AB031549  ..

 Align seq 1/1  to: AB031549  from: 1  to: 1898

        1 LeuAlaGlyLysGluLeuCysAlaMet 9
          |||||||||||||||||||||||||||
      937 CTGGCGGGCAAGGAGCTGTGCGCCATG 963

seq_name: gb_pat:AX048159

seq_documentation_block:
LOCUS        AX048159      3417 bp    DNA             PAT       15-DEC-2000
DEFINITION   Sequence 12 from Patent WO0070092.
ACCESSION    AX048159
VERSION      AX048159.1  GI:11876975
KEYWORDS
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 3417)
  AUTHORS    Kaufmann,J. and Xin,H.
  TITLE      Expression of ets domain proteins in cancer
  JOURNAL    Patent: WO 0070092-A 12 23-NOV-2000;
             Chiron Corporation (US)
FEATURES             Location/Qualifiers
     source          1..3417
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

```
                     /note "GSEF-encoding sequence with promoter (PCRs  2A-2N)"
     promoter        1. .1380
                     /note "PCR product 2A-2N"
     CDS             1815. .2822
                     /note "GSEF coding sequence"
                     /codon_start 1
                     /protein_id "CAC19604.1"
                     /db_xref="GI:11876976"
                     /translation="MGSASPCLSFVCTSHLLFPPTVCELMLEKAAACAVILEEELVS
                     PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAPGASSEEPPEEPEFQCPVIDSQAFA
                     GSLDLVPGGIEEFRSIFQVQSMVVCFVLKTQFTAVFLINHIAIPMDWSPSNVQEWLL
                     WIEHQYKLPPMGFAFLLAEELTANGIEVLKRDLLLKPHSYGKELRWLNKEK
                     TSPGATHYCASISEESWTDSEVDSSCSCQPIHLWQFLKELLLKPHSYGKELRWLNKEK
                     GIFKIEDSAQVARLWGIEEVSLAMDV
                     GIFKIEDSAQVARLWGIEEVSLAMDV
                     I"
BASE COUNT      719 a   1026 c    970 g    611 t
ORIGIN


alignment_scores:
            Quality:     9.00            Length:       9
              Ratio:    1.000              Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_178_186 x AX048159  ..

 Align seq 1/1  to: AX048159  from: 1  to: 3317

       1 LeuAlaGlyLysGluLeuCysAlaMet 9
         |||||||||||||||||||||||||||
    2346 CTGGCGGGCAAGGAGCTGTGCGCCATG 2372

seq_name: gb_pr6:AL157372

seq_documentation_block:
LOCUS           AL157372    150660 bp    DNA             PRI       29-NOV-2000
DEFINITION  Human DNA sequence from clone RP11-375E1 on chromosome 6, complete
            sequence.
ACCESSION   AL157372
VERSION     AL157372.18  GI:11493236
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 150660)
  AUTHORS   Skuce,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 30, 2000 this sequence version replaced gi:11342772.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to Sequence Map Criteria
            as follows.  An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements.  Where the sequence is
            ambiguous, there is an annotation using the 'unsure' feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            EMBL: EMBL, SWISSPROT: TrEMBL: WP:  WORMPEP: Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep.shtml
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-375E1 is from the library RPCI-11.2 constructed at the Roswell
            Park Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/.
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-375E1 it may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP11-375E1 is at 1 in this sequence. The
            true left end of clone RP3-391022 is at 150561 in this sequence.
            The true right end of clone RP11-402L10 is at 113044 in this
            sequence.
FEATURES                Location/Qualifiers
     source             1. .150660
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="6"
                        /clone="RP11-375E1"
                        /clone_lib="RPCI-11.2"
BASE COUNT     35470 a   38887 c   38740 g   37563 t
ORIGIN


alignment_scores:
            Quality:     9.00            Length:       9
              Ratio:    1.000              Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_178_186 x AL157372/rev  ..

 Align seq 1/1  to reverse of: AL157372  from: 1  to: 150660

       1 LeuAlaGlyLysGlnLeuCysAlaMet 9
         |||||||||||||||||||||||||||
  126371 CTGGCGGGCAAGGAGCTGTGCGCCATG 126345

seq_name: gb_bx3:PSEPCHRA

seq_documentation_block:
LOCUS       PSEPCHRA     1753 bp    DNA             BCT       01-SEP-1994
DEFINITION  Pseudomonas aeruginosa pchR gene, complete cds.
ACCESSION   L11657
VERSION     L11657.1  GI:388338
KEYWORDS    pchR gene.
SOURCE      Pseudomonas aeruginosa (strain PAO1) DNA.
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Pseudomonas.
REFERENCE   1  (bases 1 to 1753)
  AUTHORS   Heinrichs,D.E. and Poole,K.
  TITLE     Cloning and sequence analysis of a gene (pchR) encoding an AraC
            family activator of pyochelin and ferri-pyochelin receptor
            synthesis in Pseudomonas aeruginosa
  JOURNAL   J. Bacteriol. 175, 5882-5889 (1993)
  MEDLINE   93388519
FEATURES                Location/Qualifiers
     source             1. .1753
                        /organism="Pseudomonas aeruginosa"
                        /strain="PAO1"
                        /db_xref="taxon:287"
     -35_signal         60. .65
     -10_signal         84. .89
     RBS                254. .257
     gene               265. .1155
                        /gene="pchR"
     CDS                265. .1155
                        /gene="pchR"
                        /standard_name="PchR"
                        /codon_start=1
                        /transl_table=11
                        /function="Positive activator of pyochelin and
                        ferripyochelin receptor"
                        /evidence=experimental
```

```
                     /protein_id "AAA25926.1"
                     /db_xref "GI:388439"
                     /translation "MCIFFIAFLADAAALAEGNEFVARELENMFEVSIFCGASFD
                     WFEEPLERGLRLILVQSGQLRCRIPGGPEHLIEGPSIATIANDGDFISAQIYGHEFL
                     RYFIVQLGVEALDSRLGWLPEQLIRRPGCDPRIMSCPAPRAMQALASQIATCQMLGP
                     RDLYIGGKALELAALSAQFLSCEGPPVEEPRTPCSEVERIHAARDLLVGALQEPPSLD
                     TLASRVGMNPRKLIAGFRKVFGASVFCYLQEYRLREAHRMLCDELEANVSIVAYKVGYS
                     PAHFSIAFRKRYGISPSEIR"
BASE COUNT       275 a    616 c    571 g    291 t
ORIGIN


alignment scores:
            Quality:     8.00            Length:       8
              Ratio:    1.000              Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2_COPY_178_186 x PSEPCHRA/1--

 Align seq 1/1  to reverse of: PSEPCHRA  from 1  to 1754

        1 LeuAlaGlyLysGluLeuCysAla 8
          ||||||||||||||||||||||||
      813 CTCGGCCGGAAAGGAACTGTGCGCT 790

seq name: qb bal:AE004840

seq documentation block:
LOCUS        AE004840     10481 bp    DNA             BCT       30-AUG-2000
DEFINITION   Pseudomonas aeruginosa PA01, section 401 of 529 of the complete
             genome.
ACCESSION    AE004840 AE004091
VERSION      AE004840.1  GI:9950442
KEYWORDS     .
SOURCE       Pseudomonas aeruginosa.
  ORGANISM   Pseudomonas aeruginosa
             Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
             Pseudomonas
REFERENCE    1  (bases 1 to 10481)
  AUTHORS    Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
             Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M.,
             Garber,R.L., Goltry,L., Tolentino,E., Westbrock-Wadman,S., Yuan,Y.,
             Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
             Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
  TITLE      Complete genome sequence of Pseudomonas aeruginosa PA01, an
             opportunistic pathogen
  JOURNAL    Nature 406 (6799), 959-964 (2000)
  MEDLINE    20437347
REFERENCE    2  (bases 1 to 10481)
  AUTHORS    Stover,C.K., Pham,X.Q.T., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
             Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
             Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
             Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
             Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
             Wong,G.K.S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H.,
             Hancock,R.E.W., Lory,S. and Olson,M.V.
  TITLE      Direct Submission
  JOURNAL    Submitted (16-MAY-2000) Department of Medicine and Genetics,
             University of Washington Genome Center, University of Washington,
             Box 352145, Seattle, WA 98195, USA
FEATURES             Location/Qualifiers
     source          1..10481
                     /organism "Pseudomonas aeruginosa"
                     /strain "PA01"
                     /db_xref="taxon:287"
     gene            complement(128..1018)
                     /gene "pchR"
                     /note "PA4227"
     CDS             complement(128..1018)
                     /gene "pchR"
                     /codon_start 1
                     /transl_table 11
                     /product "transcriptional regulator PchR"
                     /protein_id "AAG07615.1"
                     /db_xref "GI:9950443"
                     /translation "MLTTLCAFPQAFAAALAFGNEFCVAFELFNMFEVSIFCGASFD
                     WFEEPLERGLRLILVQSGQLRCRIPGGPEHLIEGPSIATIANDGDFISAQIYGHEFL
                     RYFIVQLGVEALDSRLGWLPEQLIRRPGCDPRIMSCPAPRAMQALASQIATCQMLGPT
                     RDLYIGGKALELAALSAQFLSCEGPPVEEPRTPCSEVERIHAARDLLVGALQEPPSLD
                     TLASRVGMNPRKLIAGFRKVFGASVFCYLQEYRLREAHRMLCDELEANVSIVAYKVGYS
                     PAHFSIAFRKRYGISPSEIR"
gene             1247..2890
                 /gene "pchD"
                 /note "PA4228"
CDS              1247..2890
                 /gene "pchD"
                 /codon_start 1
                 /transl_table 11
                 /product "pyochelin biosynthesis protein PchD"
                 /protein_id "AAG07616.1"
                 /db_xref "GI:9950444"
                 /translation "MTSSPVTPSAVDDAFDWFAAEVRKYLDAGHWQIQSFAEALAISA
                 AEHFRRIALCDDDQRLSYADLLQRCRRLAAGLRQAGLAHGDIVVLHLPNCIAFVETCF
                 ALFQLGVRPVLALPAHRQHEISGFCRFAEAKAYIGAERIDGFDPRPMARELLASGACR
                 MALIHGEAEAPLQALAPLYQADALEDCAARAEDIACPQLSGCTTGTPKLIPRRHREYL
                 YNVRASAEVCGFDEHTVYLPGLPMAHNFTLSCPGVLGTLLAGGRVVVVSQKADPEEIC
                 LIAREKVTHTALVPPLAMLWLDAQESKRADLSSLRLLQVGGSKLGSSAAQKVEPVLGC
                 QLQQVLGMAEGLICYTRLDDPPERVLHTQGRPLSPDDEVKVVDAEGREVGPGEVGELL
                 VRGPYTIRGYYKLPEHNAKAFSADGFYRTGDRVSKDKDGYLVVEGRSKDQINRGGEKI
                 AAEEVENLLIAHPQVHDATVVAMPDSLLGERTCAFVIPRQPADPSALELKQYLHACGLA
                 AFKVPDRIELVPAFPQTGTGKISKKDLREELRRELEARA"
gene             2887..3642
                 /gene "pchC"
                 /note "PA4229"
CDS              2887..3642
                 /gene "pchC"
                 /codon_start 1
                 /transl_table 11
                 /product="pyochelin biosynthetic protein PchC"
                 /protein_id "AAG07617.1"
                 /db_xref "GI:9950445"
                 /translation "MSAAWVRPERLTPMPERLRLACPPHAGGSASEFFRSWSEKLPPDID
                 LLAIAYPGREDFNEAPATRLEDLADGAALALRDEADAPLALFGHSIGAALAYEIALR
                 LESAGAPLRHLFVSAHPAPHRQRGGALHRGDEAALLEDVRRQGGASELLEDADLRALF
                 LPILRADYQAIETYRRAQPIALACALDVLIGEDDEVSAAEAQAWSDASKIFARLRRF
                 PGGHEYLSEGRDAVIEHLLRRLAHPDALSREVA"
gene             3642..3947
                 /gene "pchB"
                 /note "PA4230"
CDS              3642..3947
                 /gene "pchB"
                 /codon_start 1
                 /transl_table 11
                 /product "salicylate biosynthesis protein PchB"
                 /protein_id "AAG07618.1"
                 /db_xref "GI:9950446"
                 /translation "MKLPRPSIGLALRERAAPSRLGLRVQAGAGERMDYVEAASELEAS
                 EAAIPAPERVAAMLPERARKWAEENGLDAPFVEGLFAQITHWYIAEQIKYWRQTHCAA"
gene             3944..5474
                 /gene "pchA"
                 /note "PA4231"
CDS              3944..5474
                 /gene "pchA"
                 /codon_start 1
                 /transl_table 11
                 /product "salicylate biosynthesis isochorismate synthase"
                 /protein_id "AAG07619.1"
                 /db_xref "GI:9950447"
                 /translation "MSRLAPLSQCLHALRGTFERATGQAQALDREVLVAASFEIDPLD
                 PLQVEGAWDDRQTPCLYWFQPELAFFAWGCALELQGHGEQREARIEENWQLLCADAVV
                 EGPLAPRLCGGFRFDPRGPREEHWQAFADASLMLAGTTVLREGERYRVLCQHLAKPGE
                 DALALAAYHCSALLRLRQPARRPSGPIAGAQGDASAQERRQWEAKVSDAVSSVRQGR
                 FGKVVLAKTGARPSGDTEFWQVIEHLRLQHADAQLLACRRGNACFLSASPERLVKTRA
                 GEALTHALAGTIARGGDAQEDARLGQALLDSAKDRHEHQLVVEAIRKTALEPFSEVLET
                 PDAPCLKKLAKVQRLNTPIRARLADAGGTLKELLQALHPTPAVGGYPEASAALDYIEQHE
```

```
                     GMDRGWYAAPLGWLDGEGNGDFLVALESALLIPGRGYLFAVNGLVGSSLFAHEYREIC
                     LKLSAMREALSAIGGLDEVPLQRGVA"
     gene            complement(5463. .5960)
                     /gene="ssb"
                     /note="PA4232"
     CDS             complement(5463. .5960)
                     /gene="ssb"
                     /codon_start=1
                     /transl_table=11
                     /product "single-stranded DNA binding protein"
                     /protein_id="AAG07620.1"
                     /db_xref="GI:9950448"
                     /translation="MARGVNKVILVGNVGSDPEVRYMPNGNAVCNISLATSESWRD
                     TGQQQERTEWHRVVFFGRLAEIAGEYLRKGSQVYVEGSLRTRKWEDQDGQDRYITEV
                     VDINGNMQMLGGRPSGDDSQRAPRPDMGRPQGAPQGGSRPAPQGQPAPQPAQDYDSFD
                     DDIPF"
     gene            complement(5977. .7365)
                     /gene="PA4233"
     CDS             complement(5977. .7365)
                     /gene="PA4233"
                     /codon_start=1
                     /transl_table=11
                     /product="probable MFS transporter"
                     /protein_id="AAG07621.1"
                     /db_xref="GI:9950449"
                     /translation "MHDSHTERMSGTETRAAVGLSLVFAFRMLGMFMVLPVLATYGQD
                     LAGATPALIGLAIGAYGLTQALLQIPFGTLSDRIGRRPVIVVGLLIFAAGAALAANAD
                     SIWGVIAGRVLQGAGAISAAVMALLSDLTPEQHRPTKAMAMIGMSIGVSFAVAMVLGPV
                     LTHLFGLHGIFWFTAGMALVGILLILFFVPQPDHMTQHRPESSVARQALLPTLKHGDLL
                     RLDAGILILHALMASFVALPLALVHEGGLPKEQHWWVYLIALLVGFFGMVPFIYAE
                     EERRMKEVLIGAVATLLACELILTLVETGSLAMLAVSIVVFIAINLGPASMPYLVGKV
                     SPAGGKGTAMGVYSTSQFLGAAIGGIIGGWMFQHGGISMVFIGCAVLAAIWLALAVIM
                     REPPYVTSIRLFLAPAAAQDAVLVQRFRALPGVALDAVVAEEAAAYVKVDIQQVIDRIA
                     LERLAQAAPGGC"
     gene            7579. .10416
                     /gene="uvrA"
                     /note="PA4234"
     CDS             7579. .10416
                     /gene="uvrA"
                     /codon_start=1
                     /transl_table=11
                     /product "excinuclease ABC subunit A"
                     /protein_id "AAG07622.1"
                     /db_xref="GI:9950450"
                     /translation "MEKLIEIAEHNLRNVNVSIPRERLVTFTGVSGAGKSSLAFDT
                     LYAEGQRRYVESLSAYARQFLSMMEKPDVDFIEGLSPAISIEQKSISHNPRSTVGTIT
                     EIYDYLRLLYARVGTPRCPDHDIPLEAQTVSQMVDQVLALPEGSKLMLLAPVIRERKG
                     EHLAVFDEMRAQGFVRARVDGKIYELDEVPKIDKQKKHSIEVVVDRFKVRADLQQRLA
                     ESFETALSIAEGIALVAFMDEDEDVEPIIPSAEFAVPVCHSISELEPKLFSFNNPAG
                     ACPTCDGLGVKQFFDARKVVNGFIIAEGAIGWDRRNVYYFQMIGSLAQHYGFSIPE
                     PFDELGAEHQKVVLYGSGRENVDFRYLNDRGDIVKRSHPFTPILPNLERRYRETESAT
                     VREELAKFLSTQPCPDCHGTRLRPEARHVWVGDPTLPAILAMPVGEAYEYAASLSLTG
                     KRCEIAAKILKEIRDRLQFLVNVGLDYLTLDRSAGTLSGGEAQRIRLASQIGAGLVGV
                     MYILDEPSIGLHQRDNERLIGTLTHLRNLGNTVIVVEHDEETIRLADYVVDIGPGAGV
                     HGGQVVAEGTPDQVMNHPDSLTGKYLSGRKKIAVPAKRRTPGNKKLILKLAGARGNNLQ
                     NVNLEIPVGLFTCITGVSGSGKSTLINNTLFPIIAIALNGATTLEVAPYDSFDGLQHL
                     DKVVDIDQSPIGRTPRSNPATYTGLFTPIRELFSGVPEARSRGYGPGRFSFNVKGGRC
                     EACQGDGVIKVEMHFLPDIYVPCDVCEGKRYNRETLEIEYEGESIHEVLEMTIEEAFE
                     FFDAVPALARKLQTLMDVGLSYIELGQSATTLSGGEAQRVKLASELHKRSTGKTLYIL
                     DEPTTGLHFADIQQLLDVLHRLRDQGNTVVVIEHNLDVIKIADWIVDLGPEGGSKGGQ
                     IIANGTPEQVAEMPQSHTGHFLKPLLERDRA"
BASE COUNT      1748 a    3715 c    3460 g    1558 t
ORIGIN


alignment_scores:
            Quality:     8.00              Length:       8
              Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_178_186 x AE004840  ..

 Align_set 1/1  to: AE004840  from: 1  to: 10481

        1 LeuAlaGlyLysGluLeuCysAla 8
          ||||||||||||||||||||||||
      470 CTCGCCGGAAAGGAACTGTGCGCT 493

seq_name: qb_bal:AF074705

seq_documentation_block:
LOCUS       AF074705     15023 bp    DNA             BCT       06-FEB-2001
DEFINITION  Pseudomonas aeruginosa dihydroaeruginoic acid synthetase (pchE),
            pyochelin synthetase (pchF), reductase (pchG), ABC transporter
            (pchH), and ABC transporter (pchI) genes, complete cds.
ACCESSION   AF074705
VERSION     AF074705.2  GI:12668017
KEYWORDS    .
SOURCE      Pseudomonas aeruginosa.
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gamma subdivision; Pseudomonadaceae;
            Pseudomonas.
REFERENCE   1  (bases 1 to 15023)
  AUTHORS   Reimmann,C., Serino,L., Beyeler,M. and Haas,D.
  TITLE     Dihydroaeruginoic acid synthetase and pyochelin synthetase,
            products of the pchEF genes, are induced by extracellular pyochelin
            in Pseudomonas aeruginosa
  JOURNAL   Microbiology 144 (Pt 11), 3135-3148 (1998)
  MEDLINE   99061207
REFERENCE   2  (bases 1 to 15023)
  AUTHORS   Reimmann,C., Patel,H.M., Serino,L., Barone,M., Walsh,C.T. and
            Haas,D.
  TITLE     Essential PchG-dependent reduction in pyochelin biosynthesis of
            Pseudomonas aeruginosa
  JOURNAL   J. Bacteriol. 183 (3), 813-820 (2001)
  MEDLINE   21142505
REFERENCE   3  (bases 1 to 15023)
  AUTHORS   Reimmann,C., Serino,L., Beyeler,M. and Haas,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-1998) Laboratoire de Biologie Microbienne,
            Universite de Lausanne, Batiment de Biologie, Lausanne, VD 1015,
            Switzerland
REFERENCE   4  (bases 1 to 15023)
  AUTHORS   Reimmann,C., Patel,H.M., Serino,L., Barone,M., Walsh,C.T. and
            Haas,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-2000) Laboratoire de Biologie Microbienne,
            Universite de Lausanne, Batiment de Biologie, Lausanne, VD 1015,
            Switzerland
  REMARK    Sequence update by submitter
COMMENT     On Feb 9, 2001 this sequence version replaced gi:3886452.
FEATURES             Location/Qualifiers
     source          1. .15023
                     /organism="Pseudomonas aeruginosa"
                     /strain="PAO1"
                     /db_xref="taxon:287"
     gene            805. .5121
                     /gene="pchE"
     CDS             805. .5121
                     /gene="pchE"
                     /note="peptide synthetase"
                     /codon_start=1
                     /transl_table=11
                     /product="dihydroaeruginoic acid synthetase"
                     /protein_id="AAC84656.1"
                     /db_xref="GI:3886453"
                     /translation="MELPPDSKTALKDWLITQLALAGLIGFFLAIVKAIAGDDDLIGVGL
                     DSIRLMYLQERLKAKGSTLDFAQLAQRPTLGAWLDLLACADRLSAPATVALPTAGRD
                     QPPELSSVQQAYWIGRGAGEVLGNVSCHAFLEFGPRDVDPQRLAAAAECVRQRHPMLR
                     ARFLDGRQQILPTPPLSCFDLQDWRTLQVDEAFRDWQALRDWRAHEGLAVERGQVFLL
                     GLVEMPGCEDRLWLSLDLLAALVESLRILAELGVAYLALELAEFFAHFADYIAHR
                     AAQRAEAAARARDYWLERLPRLPDAPALPLACAPESIRQPRTRRLAFSISAGESKRLE
                     RLAAQHGVTLSSVFGCAFALVLARWSESAEFLLNVPLFDRHADDERIGEVIADFTTLL
                     LLECRMQAGVSFAEAVKSFQRNLHGALIMAAFPALELVKEAKRQGGLESAPVVFASNL
                     GEEGFVPAAFRDAPGDLRDMLSQTPQVWLDHQLYRVGDGTLLAWDSVVGLFPEGLPET
```

```
                     MFEAYVGLLQRLCDSAWGQPADLPLPWAQQARRALLNESQPACATARTLHRGFTLRAAL
                     APDADALLYRDQRVTRCELAERALRLAGGLREAGVRPGDAVEVSLPRGPQQVAAVFGV
                     LAAGACYVPLDTDQPPARRRLIEEAAGVCLATTEEDIPQALPPRLDVQRLLRGPALAA
                     PVPLAPQASAYVIYTSGSTGVPKGVEVSHAAAINTILDALDLLLRVNASDRLLAVSAL
                     FDLSVFDLFGGLGAGASLVLPAQEQARDAAAWAEAIQRHAVSLWNSAPALLEMALSLP
                     ASQADYRSLRAVLLSGDWVALDLPGRLRPRCAEGCRLHVLGGATEAGIWSNLQEVDTV
                     PPHWRSIPYGRPLPGQAYKVVDTHGRDVPHLVVGELWIGGASLAKGYWNDPELSAKKF
                     VHDAQGRWYRTGDRGRYWGDGTLEFLGRVDQQVKVRGQRIELGEVESAALCAQACVES
                     CAAVLGGGVASLGAVLVPRLAPRAEGSMDLPAQPFAGLAEAEAVLREHLGALLEAP
                     LELDDGLRRRWLDWLADSAASALPSLDEALRRLGWQAAGLLAMGNALRGLLAGEQAPA
                     ALLLDPWLAPQAVAARLPDGREALARLLEALPTPAAGERLRVAVLDTRAGLWLDQGMA
                     SLLRPGLELTLFERSRVLLDAAATRLPERIVVQALLRGLLLFAEHAGRYDRKVISFAALH
                     AYEASREGLALAAALLRPQGRLLLVDLLCESPTALLGAALLLDRPLRLAELFSLLADL
                     AAAGLAPRCLWRSERIALVEALAFGLGLDAAALQAGLEQRLPQAMRPEPLWCLPSLPL
                     NGNGKVDRRRLAESMTRALGECCRHEPSAEEPLEAHEQALAECWEAVLKREVRRREASF
                     FSLGGDSLLATRLLAGIRERFGVRLGMADFYRQPTLAGLARHLQVQTVEEEIQLEEG
                     VL"
     gene            5118. .10547
                     /gene="pchF"
     CDS             5118. .10547
                     /gene="pchF"
                     /note="peptide synthetase"
                     /codon_start=1
                     /transl_table=11
                     /product="pyochelin synthetase"
                     /protein_id="AAC83657.1"
                     /db_xref="GI:3386454"
                     /translation="MEIGELLETRERHWAWPAWLRYRAPQGALLAGLAERLEAE
                     REALLEHLEGGPGWRAEPDMAHQRFPLTPVQAAYVLGRQAAFDYGGNACQLYAEYDWP
                     ADTDPARLEAAWNAMVERHPMLRAVIEDNAWQRVLPEVPWQRLIVHACAGLDEAAFQA
                     HLERVREELDHACAALDQWPVLRPELSIGRDACVLHCSVDELLVDYASLQLLLGEWRR
                     RYLDPQWTAEPLEATFRDYVGVEQKKPLGPAWQEFLPTWWLAFLDALPGFPLGLPLPVQP
                     DPRSTRFRHFHARLDEAAWQALGAKAGERGLSAACVALAAFAFTLGRWSQAPAFCLNL
                     TVLNRPPLHPQLAQVLGDFTALSLLAVDSRRGDSFVERARRIGEQMFDDLDHPIFSGV
                     DLLRELARRRGRGADLMPVVFTSGIGSVQRLIGDGEAPRAPRYMISQTPQVWLDCQVT
                     DQPGGLEIGWDVRIGLFTPEGQAEAMFDDFVGLLRRLAQSPRAWTLGDATEPVEAPPQA
                     LPGSARSIAAGFAERALLTPDATAIHDAAGSYSYRQVAQHASALRRVLEAHGAGRGRR
                     VAVMLPKSAAQLVAVIGILQAGAAYVPVDIRQPPLRRQALLASAEVVALVCLESDVPD
                     VGCACVAIDRLAADSAWPPPAAEVAADDLAYVIYTSGSTGTPKGVMLSHAAVSNTLL
                     DINQRYGVDANDRVLGLAELSFDLSVYDFFGATAAGAQVVLPDPARGSDPSHWAELLE
                     RHAITLWNSVPAQGQMLIDYLESEPQRHLPGPRCVLWSGDWIPVSLPTRWWRRWPDSA
                     AVQPGRRTEAAIWSIEQPTRPQHTELASIPYGRALRGQSVEVLDARGRRCPPGVRGEI
                     HIGGVGLALGYAGDPQRTAERFVRHPDGRRLYRTGDLGRYLADGSIEFLGREDDQVKI
                     RGHRIELAELDAALCAHPQVNLAATVVLGETHERSLASFVTLHAPVEAGEDPRTALDA
                     VRQRAAQALRRDWGSEEGIAAAVAALDRACLASLAAWLAGSGLFASATPLDLATLCQR
                     LGIAEARQRLLRHWLRQLEEGYLRARGEGWLGCAERPAQSPEDAWTAFAGCAPAALW
                     PAELVAYLRDSAQSLGEQLAGRISPAALMFPQGSARTAEAMYSQGLHAQALHEAMAEA
                     IAATVERQPQRRWRLLEIGACTAAASKFVIARLAPLVQRGAEVDYLFTDVSSYFLAAA
                     KERFADQPWVREGRFDMNGDLLDQGVAPHSVDTLLSSGALNNALDTPALLAGLRELLS
                     ADAWLVIQELFREHNETSVSQSLMMENPRDLRDERRQLFVHFGQWLEWLAAQGGDLAC
                     GVVPPGSALDLEGYDVLLARCKTDRARLEPAELLAFVEARVPRYMLPAQLRVLERLPV
                     TGNGKTDRKALTGFARQPQADLRHGVAQAPADELENALLALWREVLDNPSIGVEQDFF
                     GAGGDSLLIAQLIARLRERLESARRHFFDRLLRWALSQPTTPRGLAERLRSAPEEGRGP
                     ALAAARGVAPAPAGMSRAPLAEGAVALDPLVKLVPGEGVPRVLVHEGLGTLLPYRPLL
                     RALGEGRPLLGLAVHDSDAYLAIPAEHLNACLGRRYAEALHRAGLREVDLLGYCSGGL
                     VALETAKSLVQRGVRVRQLDIVSSYFTPYRVDDERLLFSFAATLGLDTAALGFPAPE
                     RLGQAVQAALAQTPERLVAEALAGLPCLADLVALRGRVLQAASGSADAVSVERDTLYR
                     LFCHSVRASQAEAPEPYVGALRLFVPDAGNPLVPRYAEALETQWRAAALGACGTHEVP
                     GGHFDCLGEALAQSLSKPMPEEASR"
     gene            10544. .11593
                     /gene="pchG"
     CDS             10544. .11593
                     /gene="pchG"
                     /note="11000"
                     /codon_start=1
                     /transl_table=11
                     /product="reductase"
                     /protein_id="AAK01463.1"
                     /db_xref="GI:12668019"
                     /translation="MSDVRGVVVAGARTGQFTAAGVAALDPETVLSGTLQGAGSKGKAL
                     AERIGVETWCEVEALPDDVRLACVAVGGAAKGEQGPALAEALMARGTDVLIEHPTLLPR
                     EWQDLLRSAERLGRRCTLLNTFYPQLPAVARFTEDGRQLHRRRGTFHLDAACGVQVGFA
                     TLDTLAALLEGVGPWSLESPSNDLSAMRGLSLVLAPVPLSLLVLNELAAADDGRMTTL
                     QKVSLTTGKTLSLGSPHGPLLWTPAVAVPAERDDQGLEALPTETAGEPLPSAQLWYAE
                     PCSWAQVHQRLWPAAAAEALALLALGDEVRRKNQRSLEVAALWQKTGERLGFPEAIPA
                     SLAPASLEQVLEQAS"
     gene            11590. .13302
                     /gene="pchH"
     CDS             11590. .13302
                     /gene="pchH"
                     /codon_start=1
                     /transl_table=11
                     /product="ABC transporter"
                     /protein_id="AAK01464.1"
                     /db_xref="GI:12668020"
                     /translation="MTPVLWRLLRTYRWRLAAAMGLQALAGLCSLLPWMLLAWLAEPL
                     AKQQAQPALLALVLLAVLAWLGCQALAAHLARRVDADLCNDLREKLLAHLQRLPLDWF
                     GRQGPDGVARLVEQDVRALHQLIAHAPNDLSNLLVVPLVALLWLAWLHPWLTLPCLLP
                     LVLAAAGFLLLRSARYRDLVLRRNAALERLSADYGEFAHNLLLARQYPGAGTQQGAEA
                     SAAAFGEAFGAWVKRVGHLAALVYVQLSTPWLLAWVLIGALALDALGVPLALAGQACAF
                     LLLLRALAAPVQALGHGGDALLGARAAAERLQQVFDQAPLAEGRSTREPVIGAVALHG
                     LGHAYEGVEVLADTDLELEDGSLVALVGPSGSGKSTLLGLLLARYMDAQRGELEVGGLA
                     LKDMDLDAVRHRHTALVGQQAAALELSLADNIALFRPDADLQEIRQAARDACLDEKIMA
                     LPRGYDSVIPGRDLQLSGGELQRLALARALLSPASLLLLDEPTSALDPQTARQVLKNLR
                     ERGCGRTRVIVAHRLAEVSDADLILVLVACGLVERGEHAALLAADGAYARLWRLQNGA
                     EVAA"
     gene            13299. .15023
                     /gene="pchI"
     CDS             13299. .15023
                     /gene="pchI"
                     /codon_start=1
                     /transl_table=11
                     /product="ABC transporter"
                     /protein_id="AAK01462.1"
                     /db_xref="GI:12668018"
                     /translation="MTLFFRMRALPPLCFAALRRASAWAVLAALLDAACGVLLVPLVE
                     AWFAEGALPRWVAALLGLSLAQALLQYLATRGFAAGGSLAAGLVKSLVARLPRLAP
                     PALRRVAPAEGLLRGPVMQAMGIPAHLIGPLIAALVTPLGVTLGLFLIDPSTAIAGLLL
                     AGAFLAALLRWSGRRNLAAEDARLAAERDAARQLQAFAERQPLLRAAQRESVARQCLE
                     EALRSLHRSTLDLLRRSLPSGLGFALAVQAAFAFALIGGAWAVERQWLDGARLVAVLV
                     LLVRFTEPLAQLTHLDQALRGAWQALDTLLRVFALAPLRSPEPGERPHDASLAAEAVE
                     LRLEDGRALLEDISLRLEPGSLNVLVGPSGAGKSSLLALLGRLYDVDAGRVLLGGVDI
                     RRLSETTLAASRNLVFQDNGLFPGSVAWNLRMARADADLEAAARAVGLLEETFAW
                     PQGWDSDVGPGGALLSGGQRQRLCLARGLLSTAPLLLLDEPTASLDAASEAQVLRSLL
                     GLRGRRTLLVLTHRPALARQADQVLLLEEGRLRLSGLHADLLVRDDWYAGFVGLAGEE
                     SSAIVVDR"
BASE COUNT      1842 a    5432 c    5443 g    2406 t
ORIGIN


alignment scores:
            quality:    8.00              length:      8
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_178_186 x AF074705/rev   ..

 Align seq 1/1  to reverse of: AF074705  from: 1  to: 15023

       1 LeuAlaGlyLysGluLeuCysAla 8
         |||||||||||||||||||||||||
     269 CTCGCCGGAAAGGAACTGTGCGCT 246

seq name: gb bt:23:AP000652

Seq documentation block:
LOCUS       AP000652    124808 bp    DNA              HTG       26-JAN-2001
DEFINITION  Homo sapiens chromosome 11 clone CMB9-865 map 11q14, WORKING DRAFT
            SEQUENCE, 16 unordered pieces.
ACCESSION   AP000652
VERSION     AP000652.3  GI:12539447
KEYWORDS    HTC; HTGS PHASE1; HTGS DRAFT.
SOURCE      Homo Sapiens DNA, clone:CMB9-865.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 124808)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 124,808 genomic DNA of 11q14
  JOURNAL   Published only in DataBase (1999) In press
REFERENCE   2  (bases 1 to 124808)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-OCT-1999) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hatt r[illegible] jp, [illegible]
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On Jan 25, 2001 this sequence version replaced gi:8118844.
            ----- - ------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://[illegible]
                Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
                Center project name: Humbrattl1
                Center clone name: CMB9-865
            ----- ------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.960429
                Consensus quality: 120416 bases at least Q40
                Consensus quality: 122044 bases at least Q30
                Consensus quality: 122878 bases at least Q20
                Insert size: 123308; sum of contigs
                Quality coverage: 10.53x in Q20 bases, Sum of contigs
            --------------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            16 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved
                  1      25828 contig of   25828 bp in length
              25929      45588 contig of   19660 bp in length
              45689      60840 contig of   15152 bp in length
              60941      75448 contig of   14508 bp in length
              75549      89149 contig of   13601 bp in length
              89250     100275 contig of   11026 bp in length
             100376     108456 contig of    8081 bp in length
             108557     114044 contig of    5488 bp in length
             114145     116068 contig of    1924 bp in length
             116169     117199 contig of    1031 bp in length
             117300     118399 contig of    1100 bp in length
             118500     119500 contig of    1001 bp in length
             119601     120778 contig of    1178 bp in length
             120879     122059 contig of    1181 bp in length
             122160     123261 contig of    1102 bp in length
             123362     124808 contig of    1447 bp in length.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 16 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    25828: contig of 25828 bp in length
            *   25829 25928: gap of      100 bp
            *   25929    45588: contig of 19660 bp in length
            *   45589 45688: gap of      100 bp
            *   45689    60840: contig of 15152 bp in length
            *   60841 60940: gap of      100 bp
            *   60941    75448: contig of 14508 bp in length
            *   75449 75548: gap of      100 bp
            *   75549    89149: contig of 13601 bp in length
            *   89150 89249: gap of      100 bp
            *   89250    100275: contig of 11026 bp in length
            *  100276 100475: gap of      100 bp
            *  100376   108456: contig of 8081 bp in length
            *  108457 108556: gap of      100 bp
            *  108557   114044: contig of 5488 bp in length
            *  114045 114144: gap of      100 bp
            *  114145   116068: contig of 1924 bp in length
            *  116069 116168: gap of      100 bp
            *  116169   117199: contig of 1031 bp in length
            *  117200 117299: gap of      100 bp
            *  117300   118399: contig of 1100 bp in length
            *  118400 118499: gap of      100 bp
            *  118500   119500: contig of 1001 bp in length
            *  119501 119600: gap of      100 bp
            *  119601   120778: contig of 1178 bp in length
            *  120779 120878: gap of      100 bp
            *  120879   122059: contig of 1181 bp in length
            *  122060 122159: gap of      100 bp
            *  122160   123261: contig of 1102 bp in length
            *  123262 123361: gap of      100 bp
            *  123362   124808: contig of 1447 bp in length.
FEATURES             Location/Qualifiers
     source          1..124808
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q14"
                     /clone="CMB9-865"
     misc_feature    1..25828
                     /note "assembly_fragment"
     misc_feature    25929..45588
                     /note "assembly_fragment"
     misc_feature    45689..60840
                     /note "assembly_fragment clone_end:17 vector_side:right"
     misc_feature    60941..75448
                     /note "assembly_fragment"
     misc_feature    75549..89149
                     /note "assembly_fragment"
     misc_feature    89250..100275
                     /note "assembly_fragment"
     misc_feature    100376..108456
                     /note "assembly_fragment"
     misc_feature    108557..114044
                     /note "assembly_fragment"
     misc_feature    114145..116068
                     /note "assembly_fragment"
     misc_feature    116169..117199
                     /note "assembly_fragment"
     misc_feature    117300..118399
                     /note "assembly_fragment"
     misc_feature    118500..119500
                     /note="assembly_fragment"
     misc_feature    119601..120778
                     /note "assembly_fragment"
     misc_feature    120879..122059
                     /note "assembly_fragment"
     misc_feature    122160..123261
                     /note "assembly_fragment"
     misc_feature    123362..124808
                     /note "assembly_fragment"
BASE COUNT     35848 a  24782 c  24562 g  38196 t   1500 others
ORIGIN


alignment_scores:
                Quality:     8.00              Length:       8
                  Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945b-2_COPY_178_186 x AP000652   ..
```

```
 Align seq 1/1  to: AP000652  from: 1  to: 124808

       2 AlaGlyLysGluLeuCysAlaMet 9
         ||||||||||||||||||||||||||
   13828 GCTGGGAAAGAACTCTGCGCAATG 13851

seq name: qb btq24:AP001876

Seq documentation block:
LOCUS       AP001876    194701 bp    DNA              HTG       03 NOV 2000
DEFINITION  Homo sapiens chromosome 11 clone CTD 2544I7 map 11q, WORKING DRAFT
            SEQUENCE, 11 unordered pieces.
ACCESSION   AP001876
VERSION     AP001876.3  GI:11094165
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens DNA, clone:CTD 2544I7.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 194701)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong Seoq,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 194,701 genomic DNA of 11q
  JOURNAL   Published only in DataBase (2000) In press
REFERENCE   2  (bases 1 to 194701)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong Seoq,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (24 APR-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7 22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, "FL5#5$,"Baq i- 5355, ge 28,
            Tel:81 45 503-9111, Fax:81 45 503 9170)
COMMENT     On Nov 3, 2000 this sequence version replaced gi:8117527.
                - - Genome Center
            Center: RIKEN Genomic Sciences Center(GSC)
            Center code: RIKEN
            Web site: http://hgp.gsc.riken.go.jp/
            Contact: hattori@gsc.riken.go.jp
                - - - Project Information
            Center project name: Humdraft11
            Center clone name: CTD 2544I7
                - - - - - Summary Statistics
            Sequencing vector: PCR products; 100% of reads
            Chemistry: Dye terminator ET amersham; 100% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 191273 bases at least Q40
            Consensus quality: 192921 bases at least Q30
            Consensus quality: 193469 bases at least Q20
            Insert size: 194701; sum of contigs
            Quality coverage: 8.77x in Q20 bases; sum of contigs
                - - -
            NOTE: This is a 'working draft' sequence. It currently consists of
            11 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved.
                  1      51760 contig of   51760 bp in length
              51861      84234 contig of   32374 bp in length
              84335     109011 contig of   24677 bp in length
             109112     134148 contig of   25037 bp in length
             134249     153108 contig of   18860 bp in length
             153209     160968 contig of    7760 bp in length
             161069     169491 contig of    8423 bp in length
             169492     178004 contig of    8513 bp in length
             178105     185843 contig of    7739 bp in length
             185944     192574 contig of    6631 bp in length
             192675     194701 contig of    2027 bp in length.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1      51760: contig of   51760 bp in length
            *   51761   51860: gap of        100 bp
            *   51861      84234: contig of   32374 bp in length
            *   84235   84334: gap of        100 bp
            *   84335     109011: contig of   24677 bp in length
            *  109012  109111: gap of        100 bp
            *  109112     134148: contig of   25037 bp in length
            *  134149  134248: gap of        100 bp
            *  134249     153108: contig of   18860 bp in length
            *  153109  153208: gap of        100 bp
            *  153209     160968: contig of    7760 bp in length
            *  160969  161068: gap of        100 bp
            *  161069     169491: contig of    8423 bp in length
            *  169492  169591: gap of        100 bp
            *  169592     178004: contig of    8513 bp in length
            *  178005  178104: gap of        100 bp
            *  178105     185843: contig of    7739 bp in length
            *  185844  185943: gap of        100 bp
            *  185944     192574: contig of    6631 bp in length
            *  192575  192674: gap of        100 bp
            *  192675     194701: contig of    2027 bp in length.
FEATURES             Location/Qualifiers
     source          1..194701
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="CTD 2544I7"
     misc_feature    1..51760
                     /note="assembly_fragment"
     misc_feature    51861..84234
                     /note="assembly_fragment"
     misc_feature    84335..109011
                     /note="assembly_fragment"
     misc_feature    109112..134148
                     /note="assembly_fragment"
     misc_feature    134249..153108
                     /note="assembly_fragment"
     misc_feature    153209..160968
                     /note="assembly_fragment"
     misc_feature    161069..169491
                     /note="assembly_fragment"
     misc_feature    169492..178004
                     /note="assembly_fragment"
     misc_feature    178105..185843
                     /note="assembly_fragment"
     misc_feature    185944..192574
                     /note="assembly_fragment"
     misc_feature    192675..194701
                     /note="assembly_fragment"
BASE COUNT    58547 a   39418 c   40037 g   55709 t    1000 others
ORIGIN


alignment scores:
            Quality:    8.00              Length:       8
              Ratio:   1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2 COPY_178_186 x AP001876  ..

 Align seq 1/1  to: AP001876  from: 1  to: 194701

       2 AlaGlyLysGluLeuCysAlaMet 9
         ||||||||||||||||||||||||||
  122151 GCTGGGAAAGAACTCTGCGCAATG 122174
```

```
seq_name: gb_pr9:HSCALRT3

seq_documentation_block:
LOCUS       HSCALRT3      260 bp    DNA             PRI       [illegible] NOV 1999
DEFINITION  Human DNA for calretinin exon 3
ACCESSION   X56670
VERSION     X56670.1  GI:29684
KEYWORDS    calcium binding protein; calretinin.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 260)
  AUTHORS   Parmentier,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV 1990) M. Parmentier, [illegible]
            ERASME, 808 ROUTE DE LENNIK, 1070 BRUSSELS, BELGIUM
REFERENCE   2  (bases 1 to 260)
  AUTHORS   Parmentier,M. and Lefort,A.
  TITLE     Structure of the human brain calcium-binding protein calretinin and
            its expression in bacteria
  JOURNAL   Eur. J. Biochem. 196 (1), 79-85 (1991)
  MEDLINE   91160569
FEATURES             Location/Qualifiers
     source          1..260
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
     intron          <1..54
                     /number 2
     exon            55..144
                     /number 3
                     /usedin X56668:calretinin_cds
                     /usedin X56668:calretinin_mRNA
     intron          145..>260
                     /number 3
BASE COUNT       71 a     57 c     78 g     54 t
ORIGIN


alignment_scores:
            Quality:    7.00            Length:       7
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_178_186 x HSCALRT3  ..

 Align seq 1/1  to: HSCALRT3  from: 1  to: 260

       1 LeuAlaGlyLysGluLeuCys 7
         |||||||||||||||||||||
     156 CTCGCTGGTAAAGAGCTGTGT 176

seq_name: gb_ro1:AB019436

seq_documentation_block:
LOCUS       AB019436     1704 bp    mRNA            ROD       20-JAN-2000
DEFINITION  Mus musculus mRNA for Prostate specific transcription factor ets,
            complete cds.
ACCESSION   AB019436
VERSION     AB019436.1  GI:4760536
KEYWORDS    Prostate specific transcription factor ets; mPse.
SOURCE      Mus musculus (strain:C57BL/6) Adult male Colon Epithelial cDNA to
            mRNA.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Yamada,N., Tamai,Y., Miyamoto,H. and Nozaki,M.
  TITLE     Cloning and expression of the mouse Pse gene encoding a novel Ets
            family member
  JOURNAL   Gene 241 (2), 267-274 (2000)
  MEDLINE   20137504
REFERENCE   2  (bases 1 to 1704)
  AUTHORS   Nozaki,M.
  TITLE     Direct Submission
  JOURNAL   [illegible] to the DDBJ/EMBL/GenBank databases. Masami
            Nozaki, Osaka University, Institute for Microbial Diseases, 3-1,
            Yamadaoka, Suita, Osaka 565-0871, Japan
            (E-mail:[illegible]biken.osaka-u.ac.jp, Tel:81-6-879-8289,
            Fax:81-6-879-8299)
FEATURES             Location/Qualifiers
     source          1..1704
                     /organism "Mus musculus"
                     /strain "C57BL/6"
                     /db_xref "taxon:10090"
                     /cell_type "Epithelial"
                     /dev_stage "Adult"
                     /sex "male"
                     /tissue_type "Colon"
     gene            362..1339
                     /gene "mPse"
     CDS             362..1339
                     /gene "mPse"
                     /codon_start 1
                     /product "prostate specific transcription factor ets"
                     /protein_id "BAA77429.1"
                     /db_xref "GI:4760537"
                     /translation "MGSASPSLSNVSFSKTLLLEPEVAPSSLELSAASSAMGLPLKGFWS
                     ESPFAIPPAQGLSAEYESYENMYPDISWVAKVPEARAGEGHPGEPGQLEYLCSASGS
                     IIDEHSLEQVQSMVVGEVLKDTEIACKELNNTIADGDWSFSGNVSKWEIWFEHQYKI
                     ASSAFQELCVKELCAMSEEFEQKAPSSQVLEAHLEKWESAAWMLKISPSLLDH A
                     SISEEGWIDGEVDSSCSGQPTHLWQFLKELLLKPHSYGEFTEWLNKEKGLIPKTEDSAQ
                     VAMLWGVPKNKPAMNYDKLSRSLPQYYKKCTERKPDISQRLVVQFVHPV"
     polyA_site      1704
                     /note="10 a nucleotides"
BASE COUNT       369 a    527 c    491 g    317 t
ORIGIN


alignment_scores:
            Quality:    7.00            Length:       7
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_178_186 x AB019436  ..

 Align seq 1/1  to: AB019436  from: 1  to: 1704

       1 GlyLysGluLeuCysAlaMet 9
         |||.|||||||||||||||||
     869 GGIAAGGAGCTGTGCGCCATG 889

seq_name: gb_pr6:AL445224

seq_documentation_block:
LOCUS       AL445224     9717 bp    DNA             PRI       08-DEC-2000
DEFINITION  Human DNA sequence from clone RP11-259P20 on chromosome 6, complete
            sequence.
ACCESSION   AL445224
VERSION     AL445224.8  GI:11611394
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 9717)
  AUTHORS   Johnson,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Dec 9, 2000 this sequence version replaced gi:11611000.
            Draft sequence assembly data is computed from overlapping clones.
```

```
            Where differences are found these are annotated as variation,
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows.  An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements.  Where the sequence is
            ambiguous, there is an annotation using the 'unsure' feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep.shtml  sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RPI1 259P20 is from the library RPCI 11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RPI1 259P20 it may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP3-493F7 is at 9618 in this sequence.
            The true right end of clone RPI 98H18 is at 100 in this sequence.
FEATURES             Location/Qualifiers
     source          1..9717
                     /organism "Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome "6"
                     /clone "RPI1 259P20"
                     /clone_lib "RPCI 11.1"
     misc_feature    1..100
                     /note "match: STS: Em:HS918H81"
     misc_feature    1..109
                     /note "match: GSS: Em:B33707"
     repeat_region   42..79
                     /note "24 copies 2 mer qa 85% conserved"
     repeat_region   245..484
                     /note "L1MA3 repeat: matches 6054 ..6298 of consensus"
     repeat_region   484..795
                     /note "AluJo repeat: matches 1..292 of consensus"
     repeat_region   796..1676
                     /note "L1MA3 repeat: matches 5204..6056 of consensus"
     misc_feature    2621..3006
                     /note "match: GSS: Em:AQ476205"
     repeat_region   4232..5467
                     /note "32 repeat: matches 1255 ..2481 of consensus"
     repeat_region   5515..5643
                     /note "L2 repeat: matches 2967..3219 of consensus"
     repeat_region   5660..5742
                     /note "MLLIF repeat: matches 9..93 of consensus"
     repeat_region   5772..6089
                     /note "MLTIF repeat: matches 242..541 of consensus"
     misc_feature    5806..6249
                     /note "match: GSS: Em:B33624"
     repeat_region   6442..6527
                     /note "MIR repeat: matches 2..88 of consensus"
     repeat_region   6565..9717
                     /note "L1P repeat: matches 11..4140 of consensus"
BASE COUNT      3023 a   2048 c   1758 g   2898 t
ORIGIN


alignment_scores:
            Quality:    7.00              Length:       7
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:

 BS:89 126 9158 2 C FY 178 186 x AL445224/rev  ..

 Align seq 1/2  to reverse of: AL445224  from: 1  to: 9717

        1 ImAlaGlyLysGluAspCys 7
          |||| ||||||||||||||||
     7691 CTTGCTGGCAAAGAGTTGTGT 7671

seq_name: ab gi|AB041007

seq_documentation_block:
LOCUS        AB041007     17063 bp    DNA              PRI       06-SEP-2000
DEFINITION   Homo sapiens DNA, MHC class I region, 7.1 ancestral haplotype.
ACCESSION    AB041007
VERSION      AB041007.2  GI:7209542
KEYWORDS     .
SOURCE       Homo sapiens (haplotype:MHC 7.1 Ancestral Haplotype HLA A3 HLA B7
             HLA DR15) Adult B Lymphoblastoid DNA.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (sites)
  AUTHORS    Gaudieri,S., Leelayuwat,C., Townend,D.C., Kulski,J.K. and
             Hawkins,B.R.
  TITLE      Genomic characterization of the region between HLA B and TNF:
             implications for the evolution of multicopy gene families
  JOURNAL    J. Mol. Evol. 44 Suppl 1, S147-S154 (1997)
  MEDLINE    97224649
REFERENCE    2  (bases 1 to 17063)
  AUTHORS    Hawkins,B.R. and Gaudieri,S.
  TITLE      Direct Submission
  JOURNAL    Submitted (03-AUG-1999) Prof B.R. Hawkins, Molecular
             Immunology and Instrumentation Laboratories, School of Biomedical
             Science, University of Western Australia, Nedlands, WA 6907,
             Australia (E-mail:sgaudier@labs.uwa.edu.au, Tel:61 8 93463414,
             Fax:61 8 93464578)
COMMENT      Sequence incomplete and contains sequence with only single strand
             coverage.
             Sequence contains some ambiguities.
FEATURES             Location/Qualifiers
     source          1..17063
                     /organism "Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_type "B Lymphoblastoid"
                     /chromosome "6"
                     /dev_stage "Adult"
                     /haplotype "MHC 7.1 Ancestral Haplotype HLA A3 HLA B7
                     HLA-DR15"
                     /map "6p21.3"
     repeat_unit     1831..1952
                     /rpt_family "AluJo,FLAM"
     repeat_unit     complement(2076..2497)
                     /rpt_family "LIR33"
     repeat_unit     2696..2720
                     /note "AT rich"
     repeat_unit     3154..3489
                     /rpt_family "THE1A"
     repeat_region   3772..3865
                     /note "(CA)n"
     repeat_unit     4292..4441
                     /rpt_family "MLT2D"
     repeat_region   4584..4613
                     /note "(TTTC)n"
     repeat_unit     4616..5023
                     /rpt_family "Tigger1"
     repeat_unit     5011..5368
                     /rpt_family "Tigger1"
     misc_feature    6335..7513
                     /note "3.8.1 homologous region"
     misc_feature    complement(6335..7513)
                     /note "3.8.1"
     repeat_unit     6796..7416
                     /rpt_family "L2"
```

```
     repeat_unit     7424. .8005
                     /rpt_family "MLT1E2"
     repeat_unit     8037. .8115
                     /rpt_family "L2"
     repeat_unit     8218. .8307
                     /rpt_family "L2"
     repeat_unit     8423. .9149
                     /rpt_family "L2"
     repeat_unit     9150. .9459
                     /rpt_family "LTR16B"
     repeat_unit     9460. .9698
                     /rpt_family "AluSx"
     repeat_unit     9699. .9861
                     /rpt_family "LTR16B"
     repeat_unit     9864. .10392
                     /rpt_family "HERV16"
     repeat_unit     10541. .10691
                     /rpt_family "HERV16"
     repeat_unit     complement(10692. .11034)
                     /rpt_family "THE1B"
     repeat_unit     11035. .12506
                     /rpt_family "HERV16"
     repeat_unit     12518. .12823
                     /rpt_family "AluSp"
     repeat_unit     12826. .12991
                     /rpt_family "AluSq/x"
     repeat_unit     12860. .13812
                     /rpt_family "HERVL"
     misc_feature    12934. .15552
                     /note "P5-1 homologous region"
     misc_feature    complement(12934. .15552)
                     /note="P5_1"
     repeat_unit     13925. .15020
                     /rpt_family "HERVL"
     repeat_unit     15039. .15308
                     /rpt_family "LTR16B"
     repeat_unit     complement(16118. .16199)
                     /rpt_family "MER4C"
     misc_feature    16253. .16898
                     /note "HLA-Xup homologous region"
     repeat_unit     complement(16719. .17055)
                     /rpt_family "MER21B"
BASE COUNT     4266 a    4278 c    4111 g    4366 t      42 others
ORIGIN


alignment_scores:
            Quality:     7.00              Length:       7
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_178_186 x AB031007/rev    ..

 Align seq 1/1  to reverse of: AB031007  from: 1  to: 17063

       2 AlaGlyLysGluLeuCysAla 8
         ||||||||||||||||||||||
    4185 GCAGGGAAGGAGCTGTGTGCT 4165

seq_name: gb_pri:AB000879

seq_documentation_block:
LOCUS        AB000879     39436 bp    DNA             PRI       20-NOV-1999
DEFINITION   Homo sapiens genomic DNA for centromeric end of MHC class I region
             on chromosome 6, cosmid clone: IY3A9, complete sequence.
ACCESSION    AB000879
VERSION      AB000879.1  GI:3021695
KEYWORDS     HTG.
SOURCE       Homo sapiens (haplotype:A2 B62 CW10 DR4) adult male blood
             immunoresponce cell cell_line:BOLETH DNA, clone_lib:CEPH YAC
             clone:745d12 sub_clone:cosmid clone:IY3A9.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (sites)
  AUTHORS    Shiina,T., Tamiya,G., Oka,A., Yamagata,T., Yamagata,N., Fukuda,F.,
             Goto,K., Miyaki,N., Watanabe,K., Fukuzumi,Y., Iwachi,S.,
             Sugawara,C., Ono,A., Chen,L., Yamazaki,M., Tashiro,H., And ,A.,
             Ikemura,T., Kimura,M. and Inoko,H.
  TITLE      Nucleotide sequencing analysis of the 146-kilobase segment around
             the IkBL and MICA genes at the centromeric end of the HLA class I
             region
  JOURNAL    Genomics 47 (3), 372-382 (1998)
  MEDLINE    98149985
REFERENCE    2  (bases 1 to 39436)
  AUTHORS    Shiina,T.
  TITLE      Direct Submission
  JOURNAL    Submitted (01-FEB-1997) to the EMBL/GenBank/DDBJ databases. Takashi
             Shiina, Tokai University School of Medicine, Department of
             Molecular Life Sciences; Bohseidai, Isehara, Kanagawa 259-11, Japan
             (E-mail:tshiina@is.icc.u-tokai.ac.jp, Tel:0463-93-1121,
             Fax:0463-94-8884)
FEATURES             Location/Qualifiers
     source          1. .39436
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /cell_line="BOLETH"
                     /cell_type "immune response cell"
                     /chromosome "6"
                     /clone "745D12"
                     /clone_lib "CEPH YAC"
                     /dev_stage="adult"
                     /haplotype "A2 B62 CW10 DR4"
                     /map "6p21.3"
                     /sex "male"
                     /sub_clone "cosmid clone:IY3A9"
                     /tissue_type "blood"
BASE COUNT     10807 a    9418 c    8837 g    10474 t
ORIGIN


alignment_scores:
            Quality:     7.00              Length:       7
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_178_186 x AB000879/rev    ..

 Align seq 1/1  to reverse of: AB000879  from: 1  to: 39436

       2 AlaGlyLysGluLeuCysAla 8
         ||||||||||||||||||||||
    2525 GCAGGGAAGGAGCTGTGTGCT 2505
```

OM olig: US-09-126-945B-2_COPY_178_186 [illegible]

Date: Sep 14, 2001 9:48 PM

About: Results were produced by the [illegible]
Copyright (c) 1993-2001 Compugen Ltd

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=[illegible]
-DB=N_Geneseq_0601 -QFMT=fastap -SUFFIX=olip2n.rng -GAPOP=1.500
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality
-THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -MINLEN=[illegible]
-MAXLEN=2000000000 -USER=US09126945_@CGN1_1_308 -NCPU=6 -ICPU=3
-LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
query: US-09-126-945B-2_COPY_178_186
query length: 9
database: N_Geneseq_0601:*
Database sequences: 730101
Database length: 313950809
Search time (sec): 680.150000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	! Documentation
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF21828 +	9.00	178.78	0.0548	1087	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ52[illegible] +	9.00	174.90	0.0902	1894	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:AAC83[illegible] +	[illegible]	17[illegible]	[illegible]	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:AAC83266 +	9.00	170.99	0.1490	3317	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AA[illegible] +	[illegible]	[illegible]	[illegible]	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:AAX[illegible] +	[illegible]	127.13	41.[illegible]	2[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV68325 +	6.00	127.13	41.29	288	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AA[illegible] +	[illegible]	127.13	41.[illegible]	2[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA[illegible] +	[illegible]	[illegible]	[illegible]	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA[illegible] +	[illegible]	[illegible]	[illegible]	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AA[illegible] +	6.00	120.53	[illegible]	711	!
/SIDS1/gcgdata/geneseq/geneseqn/NA[illegible] +	[illegible]	[illegible]	[illegible]	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA[illegible] +	6.00	118.[illegible]	125.[illegible]	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ[illegible] +	6.00	115.07	193.91	1620	!
/SIDS1/gcgdata/geneseq/geneseqn/NA[illegible] +	[illegible]	[illegible]	[illegible]	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AA[illegible] +	6.00	109.65	[illegible]	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AA[illegible] +	[illegible]	[illegible]	[illegible]	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV[illegible] +	[illegible]	[illegible]	[illegible]	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AA[illegible] +	[illegible]	102.82	923.22	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF[illegible]405 +	6.00	102.82	923.22	9465	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA[illegible] +	[illegible]	[illegible]	[illegible]	11156	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF[illegible] +	6.00	[illegible]	[illegible]	11747	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF[illegible] +	6.00	[illegible]	[illegible]	15[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF97854 +	[illegible]	[illegible]	[illegible]	21490	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AA[illegible] +	[illegible]	81.[illegible]	[illegible]	121162	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX[illegible] +	6.00	68.76	7.2e+04	129025	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ[illegible] +	[illegible]	126.19	46.62	18	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ[illegible] +	[illegible]	125.15	51.23	2[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV4[illegible] +	6.00	124.18	60.72	24	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAX82692 +	6.00	124.18	60.72	24	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX21747	[illegible]	123.62	64.91	26	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAF[illegible] +	[illegible]	122.62	73.67	30	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAB[illegible] +	[illegible]	122.62	73.67	30	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA[illegible] +	[illegible]	121.65	[illegible]	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV59868 +	[illegible]	120.97	91.03	39	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV59869 +	6.00	120.97	91.03	39	!
/SIDS1/gcgdata/geneseq/geneseqn/NA1991.DAT:AA[illegible] +	[illegible]	[illegible]	[illegible]	[illegible]	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AA[illegible] +	[illegible]	118.02	1[illegible]	59	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC15545 +	5.00	117.44	143.16	63	!
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC[illegible] +	[illegible]	[illegible]	[illegible]	106	
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC25060 +	5.00	112.49	270.11	128	
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC11911 +	5.00	112.38	273.88	130	
/SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV4[illegible] +	[illegible]	111.47	[illegible]	148	
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAX82643 +	5.00	111.47	307.61	148	
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC15791 +	5.00	111.24	316.90	153	

seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF21828

seq_documentation_block:
ID AAF21828 standard; DNA; 1087 BP.
XX
AC AAF21828;
XX
DT 27-MAR-2001 (first entry)
XX
DE Human breast and ovarian cancer associated antigen gene SEQ ID 215.
XX
KW Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
KW nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
KW antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
KW Addison's disease; allergy; autoimmune haemolytic anaemia;
KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW cardiovascular disorder; wound healing; neurological disease; ds.
XX
OS Homo sapiens.
XX
PN WO200055173-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05881.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI [illegible] A, [illegible] SM;
XX
DR WPI; [illegible]
DR P-PSDB; AAB58925.
XX
PT New human breast and ovarian cancer associated gene sequences and the
PT polypeptides encoded by these genes, useful in the prevention,
PT treatment and diagnosis of cancer, immune disorders, cardiovascular
PT disorders and neurological diseases -
XX
PS Claim [illegible]; Page 642-643; 1209pp; English.
XX
CC Sequences AAF21[illegible] - AAF22041 represent DNA sequences encoding human
CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
CC associated with breast and ovarian cancer. Included in the invention are
CC sequences AAF22[illegible] - AAF22049 and AAB59129 which are used in the
CC isolation and characterisation of the DNA and protein sequences of the
CC invention. The breast and ovarian cancer associated DNA, protein, agonist
CC or antagonist sequences exhibit cytostatic; immunosuppressive;
CC nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
CC antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
CC antibacterial, antifungal, antiparasitic and cardiant activity. The
CC polynucleotide and protein sequences are used in the diagnosis of cancer,
CC particularly breast and ovarian cancer. The nucleic acid sequences,
CC proteins, agonists and agonists may also be used in the diagnosis,
CC prevention and treatment of immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis, cardiovascular disorders such as
CC myocardial ischaemia; wound healing; neurological diseases such as
CC cerebral anoxia and epilepsy; and infectious diseases.
XX
SQ Sequence 1087 BP; 215 A; 347 C; 339 G; 182 T; 4 other;

```
alignment scores:
            Quality:     9.00              Length:       9
              Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2 COPY 178-186 x AAF21828    ..

 Align seq 1/1   to: AAF21828   from: 1   to: 1087

        1 LeuAlaGlyLysGluLeuCysAlaMet 9
          ||||||||||||||||||||||||||||
      408 CTGGCGGGCAAGGAGCTGTGCGCCATG 434

seq_name: /cgn2/ftp/data/geneseq/geneseqp_NA2000_DAT.AAZ50691

seq_documentation_block:
ID   AAZ50691 standard; cDNA; 1894 BP.
XX
AC   AAZ50691;
XX
DT   23-MAY-2000  (first entry)
XX
DE   Human prostate derived Ets factor cDNA.
XX
KW   Prostate derived Ets factor; PDEF; human; chromosome 6p21.3; cancer;
KW   loss of heterozygosity; chromosomal translocation; linkage analysis;
KW   cytostatic; cardiant; immunosuppressive; cerebroprotective; fungicide;
KW   antibacterial; vulnerary; neuroprotective; antiparkinsonian; nootropic;
KW   anabolic; antiinflammatory; anorectic; hybridisation probe; forensic;
KW   tumour marker; diagnosis, treatment, prostate cancer, Ets transcription
KW   autoimmune disorder; haematopoietic; immune/nervous system; stroke;
KW   neoplasm; microbial infection; tissue regeneration; heart attack;
KW   scarring; food additive; preservative; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers:
FT   CDS             416..1423
FT                   /*tag= a
FT                   /product= "Human prostate derived Ets factor"
FT                   /note= "Expressed in prostate epithelium"
XX
PN   WO200006589-A1.
XX
PD   10-FEB-2000.
XX
PF   02-AUG-1999;   99WO-US17470.
XX
PR   31-JUL-1998;   98US-0126945.
XX
PA   (HUMA ) HUMAN GENOME SCI INC.
PA   (BETH ) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI   Libermann TA, Oettgen JP, Komsch CA, Endress GA, Rosen CA;
XX
DR   WPI; 2000-195255/17.
DR   P-PSDB; AAY45003.
XX
PT   Novel prostate derived polypeptide, polynucleotide useful for
PT   diagnosis, prevention and treatment of prostate cancer, autoimmune
PT   disorders, microbial infections and also as food additive or
PT   preservative
XX
PS   Claim 4; Fig 1; 142pp; English.
XX
CC   The present cDNA sequence encodes a human prostate derived Ets factor
CC   (PDEF), isolated from human prostate carcinoma cDNA library
CC   (ATCC No.203072). PDEF gene expression is highest in tissues involved in
CC   androgen sensitivity, with enriched expression in prostate epithelium
CC   cells. PDEF gene is mapped to the human chromosome 6p21.3
CC   region that is associated with loss of heterozygosity and chromosomal
CC   translocations in various human cancers. PDEF has cytostatic, cardiant,
CC   immunosuppressive, cerebroprotective, fungicide, antibacterial,
CC   vulnerary, neuroprotective, antiparkinsonian, nootropic, anabolic,
CC   antiinflammatory and anorectic activity. PDEF polynucleotides are useful
CC   in linkage analysis as markers, as hybridisation probes for differential
CC   identification of the tissues or cell types and as polymorphic markers
CC   for forensic purposes. PDEF is useful as prostate specific tumour marker
CC   for the diagnosis and treatment of prostate cancer. PDEF sequences are
CC   useful for treating autoimmune disorders, haematopoietic, blood
CC   coagulation, immune and nervous system disorders, hyperproliferative
CC   disorders like, neoplasms and microbial infections, heart attacks,
CC   stroke, scarring and for tissue regeneration. They are also useful as
CC   food additives or preservatives.
XX
SQ   Sequence 1894 BP; 468 A; 653 C; 571 G; 302 T; 0 other,


alignment scores:
            Quality:     9.00              Length:       9
              Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2_COPY 178-186 x AAZ50691    ..

 Align seq 1/1   to: AAZ50691   from: 1   to: 1894

        1 LeuAlaGlyLysGluLeuCysAlaMet 9
          ||||||||||||||||||||||||||||
      947 CTGGCGGGCAAGGAGCTGTGCGCCATG 973

seq_name: /cgn2/ftp/data/geneseq/geneseqp_NA2001_DAT.AAC84261

seq_documentation_block:
ID   AAC84261 standard; DNA; 1894 BP.
XX
AC   AAC84261;
XX
DT   16-MAR-2001  (first entry)
XX
DE   Gland specific Ets transcription factor (GSEF) nucleotide sequence.
XX
KW   Transcription factor; gland specific Ets transcription factor; GSEF;
KW   metastatic potential; cancer; tumour; metastasis; breast; prostate;
KW   leukaemia; lymphoma; sarcoma; melanoma; chromosome 6p21.1-6p21.3; ds.
XX
OS   Homo sapiens.
XX
PN   WO200070092-A1.
XX
PD   23-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US13173.
XX
PR   14-MAY-1999;   99US-0134112.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Kaufmann J, Xin H, Harrower G;
XX
DR   WPI; 2001-038019/05.
DR   P-PSDB; AAB49628.
XX
PT   Detecting metastatic and potential metastatic cancerous cells, useful
PT   for diagnosing, prognosing, grading and staging of cancers by detecting
PT   gland-specific Ets transcription factor gene product in a biological
PT   sample from a cell
XX
PS   Disclosure; Page 81-83; 95pp; English.
XX
CC   This invention relates to a method for the detection and determination of
CC   the metastatic potential of a cell. The method comprises detecting a
CC   gland-specific Ets transcription factor (GSEF) gene product in a test
```

```
CC   sample. Detection of a GSEF gene product in the test sample in amount
CC   lower than that in a normal cell, is indicative of a cell with high
CC   metastatic potential. The method is useful for determining the metastatic
CC   potential of a cell, for the diagnosis and prognosis of cancer as well as
CC   grading and staging of cancers by detecting GSEF expression in a
CC   biological test sample. The method may also be used to monitor patients
CC   having a predisposition to develop a particular cancer. GSEF polypeptides
CC   are useful for producing antibodies, in cancer diagnosis, prognosis,
CC   grading, staging and management of breast and prostate tumours, and in
CC   detecting polymorphisms in the sequence. GSEF genes and proteins are also
CC   useful in gene therapy. GSEF gene product expression levels can be used
CC   in conjunction with any tissue in which an alteration in GSEF gene
CC   product expression levels is associated with development of a
CC   cancer-associated phenotype. Cancers, which can be monitored include
CC   cancers of the prostate, cervix, lung and colon, melanomas, colorectal
CC   adenocarcinomas, Wilm's tumour, retinoblastoma, sarcomas, myosarcomas,
CC   lung carcinomas, leukaemia, and lymphomas. The GSEF gene is located on
CC   human chromosome 6, specifically at 6p21.1-6p21.3. The present sequence
CC   represents the DNA encoding GSEF.
XX
SQ   Sequence 1894 BP; 368 A; 653 C; 571 G; 302 T; 0 other;


alignment_scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_178_186 x AAC83261    ..

 Align seq 1/1  to: AAC83261  from: 1  to: 1894

        1 LeuAlaGlyLysGluLeuCysAlaMet 9
          |||||||||||||||||||||||||||
      947 CTGGCGGGCAAGGAGCTGTGCGCCATG 974

seq_name: /cgi2/icgi/data/geneseq/geneseqpro/NA.con.1AF AAC83266

seq_documentation_block:
ID   AAC83266 standard; DNA; 3317 BP.
XX
AC   AAC83266;
XX
DT   16-MAR-2001  (first entry)
XX
DE   Gland-specific Ets transcription factor (GSEF) cDNA sequence.
XX
KW   Transcription factor; gland-specific Ets transcription factor; GSEF;
KW   metastatic potential; cancer; tumour; metastasis; breast; prostate;
KW   leukaemia, lymphoma, sarcoma, melanoma, chromosome 6p21.1-6p21.3, ss.
XX
OS   Homo sapiens.
XX
PN   WO200070092-A1.
XX
PD   23-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US13173.
XX
PR   14-MAY-1999;   99US-0134112.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Kaufmann J,  Xin H,  Harrowe G;
XX
DR   WPI; 2001-041019/05.
DR   P-PSDB; AAB49633.
XX
PT   Detecting metastatic and potential metastatic cancerous cells, useful
PT   for diagnosing, prognosing, grading and staging of cancers by detecting
PT   gland-specific Ets transcription factor gene product in a biological
PT   sample from a cell -
XX
PS   Disclosure; Fig 2; 95pp; English.
XX
CC   This invention relates to a method for the detection and determination of
CC   the metastatic potential of a cell. The method comprises detecting a
CC   gland-specific Ets transcription factor (GSEF) gene product in a test
CC   sample. Detection of a GSEF gene product in the test sample in amount
CC   lower than that in a normal cell, is indicative of a cell with high
CC   metastatic potential. The method is useful for determining the metastatic
CC   potential of a cell, for the diagnosis and prognosis of cancer as well as
CC   grading and staging of cancers by detecting GSEF expression in a
CC   biological test sample. The method may also be used to monitor patients
CC   having a predisposition to develop a particular cancer. GSEF polypeptides
CC   are useful for producing antibodies, in cancer diagnosis, prognosis,
CC   grading, staging and management of breast and prostate tumours, and in
CC   detecting polymorphisms in the sequence. GSEF genes and proteins are also
CC   useful in gene therapy. GSEF gene product expression levels can be used
CC   in conjunction with any tissue in which an alteration in GSEF gene
CC   product expression levels is associated with development of a
CC   cancer-associated phenotype. Cancers, which can be monitored include
CC   cancers of the prostate, cervix, lung and colon, melanomas, colorectal
CC   adenocarcinomas, Wilm's tumour, retinoblastoma, sarcomas, myosarcomas,
CC   lung carcinomas, leukaemia, and lymphomas. The GSEF gene is located on
CC   human chromosome 6, specifically at 6p21.1-6p21.3. The present sequence
CC   represents the cDNA encoding GSEF.
XX
SQ   Sequence 3317 BP; 710 A; 1026 C; 970 G; 611 T; 0 other;


alignment_scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_178_186 x AAC83266    ..

 Align seq 1/1  to: AAC83266  from: 1  to: 3317

        1 LeuAlaGlyLysGluLeuCysAlaMet 9
          |||||||||||||||||||||||||||
     2346 CTGGCGGGCAAGGAGCTGTGCGCCATG 2372

Seq name: /cgi2/icgi/data/geneseq/geneseqpro/NA2000.DAT AAC09764

Seq_documentation_block:
ID   AAC09764 standard; cDNA; 215 BP.
XX
AC   AAC09764;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 13839.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
```

```
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures
XX
PS   Claim 1; SEQ ID 14839; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 40 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 215 BP; 46 A; 63 C; 73 G; 42 T; 1 other;


alignment scores:
            Quality:     6.00              Length:       6
              Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2 (COPY 178_186 x AAC09764/rc)

 Align seq 1/1  to reverse of: AAC09764  from: 1  to: 215

        8 GlyLysGluLeuCysAla 8
          ||||||||||||||||||
       71 GGYAAGGAGCTCTGTGCT 54

seq_name: /STNSL/pepdata/geneseq/geneseqp_NA1/07.DAT:AAX83387

seq_documentation block:
ID   AAX83387 standard; cDNA; 288 BP.
XX
AC   AAX83387;
XX
DT   31-AUG-1999  (first entry)
XX
DE   Breast cancer tumour specific clone #141.
XX
KW   Breast cancer; tumour; gene expression; genome; diagnosis; mammal;
KW   human endogenous retrovirus; vaccine; ss.
XX
OS   Homo sapiens.
XX
PN   WO9925426-A2.
XX
PD   17-JUL-1997.
XX
PF   10-JAN-1997;   97WO-US00485.
XX
PR   20-AUG-1996;   96US-0700014.
PR   11-JAN-1996;   96US-0585392.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Frudakis TN,  Reed SG,  Smith JM;
XX
DR   WPI; 1997-472865/44.
XX
PT   Breast cancer related DNA from retrovirus antigen (s)  useful for
PT   diagnosis and treatment of breast cancer
XX
PS   Claim 1; Page 167-168; 221pp; English.
XX
CC   Sequences AAX83201-X83285 and AAX83331-X83415 represent novel breast
CC   cancer tumour specific clones which are expressed from a genomic region
CC   containing a human endogenous retrovirus (AAX83330). Detection of
CC   the clone sequences allows determination of the presence of breast
CC   cancer in a mammal. Progression of breast cancer can be monitored by
CC   detecting the level of clone expression. Polypeptides encoded by the
CC   clones can be used in vaccines to inhibit or prevent breast cancer.
XX
SQ   Sequence 288 BP; 90 A; 61 C; 74 G; 63 T; 0 other;


alignment scores:
            Quality:     6.00              Length:       6
              Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2 (COPY 178_186 x AAX83387)  ..

 Align seq 1/1  to: AAX83387  from: 1  to: 288

        4 LysGluLeuCysAlaMet 9
          |||||||||||||||||
      217 AAGGAGCTTTGTGCCATG 234

seq_name: /STNSL/pepdata/geneseq/geneseqp_NA1/08.DAT:AAV68925

seq_documentation block:
ID   AAV68925 standard; DNA; 288 BP.
XX
AC   AAV68925;
XX
DT   22-JAN-1999  (first entry)
XX
DE   DNA molecule encoding a breast tumour specific polypeptide #117.
XX
KW   Human; breast cancer; breast tumour tissue; diagnosis; treatment;
KW   vaccine; epitope; endogenous; retroviral element; ss.
XX
OS   Homo sapiens.
XX
PN   WO9845328-A2.
XX
PD   15-OCT-1998.
XX
PF   09-APR-1998;   98WO-US06939.
XX
PR   11-DEC-1997;   97US-0991789.
PR   09-APR-1997;   97US-0838762.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Frudakis TN,  Reed SG,  Smith JM;
XX
DR   WPI; 1998-552173/47.
XX
PT   New DNA sequences isolated from endogenous human retroviral element
PT     and related vectors, transformed cells, proteins and antibodies,
PT   useful for diagnosis, treatment and prevention of breast cancer
XX
PS   Claim 1; Page 102; 178pp; English.
XX
CC   AAV68800 to AAV68998 represent nucleotide sequences which encode human
CC   breast tumour specific polypeptides. Detection or measurement of
CC   human breast tumour specific polypeptides and nucleotide sequences,
CC   or the corresponding RNA in a sample, is used for diagnosis and
CC   monitoring of breast cancer. Human breast tumour specific polypeptides
CC   and nucleotide sequences, and the vectors containing the DNAs, are also
CC   useful in vaccines for inhibiting development (for prevention or
CC   therapy) of breast cancer. The polypeptides may also be used to
CC   raise monoclonal antibodies, used as immunoassay reagents.
XX
```

```
SQ   Sequence 288 BP; 90 A; 61 C; 74 G; 63 T; 0 other;


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_178_186 x AAV68925    ..

 Align seq 1/1  to: AAV68925  from: 1  to: 288

       4 LysGluLeuCysAlaMet 9
         ||||||||||||||||||
     217 AAGGAGCTTTGTGCCATG 234

seq_name: "SIDSi":/cdata4/geneseq/published/DNA_NEW/AAV68925.rng

seq_documentation_block:
ID   AAC80910 standard; cDNA; 288 BP.
XX
AC   AAC80910;
XX
DT   13-FEB-2001  (first entry)
XX
DE   Human breast tumour-specific cDNA SEQ ID No: 198.
XX
KW   Human; breast tumour-specific antigen; cytostatic; vaccine;
KW   breast cancer; B18Aq1; B11Aq1; B15Aq1; ss.
XX
OS   Homo sapiens.
XX
PN   WO200061753-A2.
XX
PD   19-OCT-2000.
XX
PF   07-APR-2000; 2000WO-US09312.
XX
PR   09-APR-1999;   99US-0289198.
PR   28-OCT-1999;   99US-0429755.
PR   24-MAR-2000;   2000US-0534825.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Frudakis TN,  Smith JM,  Reed SG,  Misher LE,  Retter MW,  Dillon DC;
XX
DR   WPI; 2000-628403/60.
XX
PT   An isolated polypeptide comprising an immunogenic portion of a breast
PT   tumor protein used for inhibiting the development of cancer, especially
PT   breast cancer, and monitoring cancer progression in a patient
XX
PS   Claim 4; Page 142; 187pp; English.
XX
CC   The present sequence is given in a specification relating to compositions
CC   and methods for the treatment and diagnosis of breast cancer. Nucleotide
CC   sequences that are preferentially expressed in breast tumour tissue, and
CC   the polypeptides encoded by such nucleotide sequences, are used in
CC   compositions and vaccines to inhibit the development of cancer,
CC   especially breast cancer. The progression of a cancer may be monitored by
CC   carrying out detection of tumour-specific antigens at subsequent time
CC   points and comparing the results from the different time points.
CC   CD4+ and/or CD8+ T-Cells isolated from the cancer patient may be treated
CC   with tumour-specific polypeptides, polynucleotides encoding the
CC   polypeptides or antigen presenting cells expressing the polypeptides. The
CC   cells are then administered to the patient to inhibit development of
CC   cancer.
XX
SQ   Sequence 288 BP; 90 A; 61 C; 74 G; 63 T; 0 other;


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_178_186 x AAC80910    ..

 Align seq 1/1  to: AAC80910  from: 1  to: 288

       4 LysGluLeuCysAlaMet 9
         ||||||||||||||||||
     217 AAGGAGCTTTGTGCCATG 234

seq_name: "SIDSi":/cdata4/geneseq/published/DNA_NEW/AAC26222.rng

seq_documentation_block:
ID   AAC26222 standard; cDNA; 387 BP.
XX
AC   AAC26222;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 40297.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401 A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GESE ) GENSET.
XX
PI   Dumas Milne Edwards J;  Joulert A;  Giordano J;
XX
DR   WPI; 2000-581481/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures
XX
PS   Claim 1; SEQ ID 40297; 71pp + CD ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo(dT) primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 387 BP; 114 A; 79 C; 79 G; 111 T; 4 other;


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
```

```
 US-09-126-945B-2 (COPY 178_186 x AAC26222/rev   ..

 Align Seq 1/1  to reverse of: AAC26222  from: 1  to: 387

        1 LeuAlaGlyLysSilAsn 6
          ||||||||||||||||||
       54 TTAGCTGGCAAAGAGTTG 37

Seq name: /cds1/genpept/genseqpep/genseqpep_NA2000.DAT:AAV89416;

Seq documentation block:
ID   AAV89416 standard; cDNA; 513 BP.
XX
AC   AAV89416;
XX
DT   15-FEB-1999  (first entry)
XX
DE   EST clone CN516.
XX
KW   Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
KW   tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;
KW   receptor; ligand; thrombolytic; anti inflammatory; cadherin; anti tumour;
KW   gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO9845436-A2.
XX
PD   15-OCT-1998.
XX
PF   10-APR-1998;   98WO-US06955.
XX
PR   10-APR-1997;   97US-0838821.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ,  Jacobs K,  Lavallie ER,  McCoy JM,  Merberg D;
PI   Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1999-070077/06.
XX
PT   New polynucleotides encoding human secreted proteins   derived from
PT   e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT   ovary, pituitary, retina and colon cDNA libraries.
XX
PS   Claim 1; Page 211; 618pp; English.
XX
CC   The present sequence represents a human expressed sequence tag (EST).
CC   The polynucleotide, which is a secreted EST, and the encoded protein
CC   are predicted to have useful biological activities which would make
CC   them suitable for treating, preventing or ameliorating medical
CC   conditions in humans and animals, although no supporting data is
CC   given. Suggested activities include nutritional activity, immune
CC   stimulating or suppressing activity, haematopoiesis regulating
CC   activity, tissue growth activity, activin/inhibin activity,
CC   chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC   activity, receptor/ligand activity, anti inflammatory activity,
CC   cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The polynucleotide may also be useful for gene therapy.
XX
SQ   Sequence 513 BP; 167 A; 95 C; 143 G; 106 T; 2 other;


alignment scores:
             Quality:     6.00              Length:       6
               Ratio:    1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2 (COPY_178_186 x AAV89416   ..

 Align Seq 1/1  to: AAV89416  from: 1  to: 513

        4 LysGlnLeuCysAlaMet 9
          ||||||||||||||||||
      127 AAGCAGCTTGCTACATG 144

Seq name: /cds1/genpept/genseqpep/genseqpep_NA2000.DAT:AAC40930

Seq_documentation block:
ID   AAC40930 standard; DNA; 741 BP.
XX
AC   AAC40930;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 80042.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134470.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
```

PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149724.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151403.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152463.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154089.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155149.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159429.
PR 14-OCT-1999; 99US-0159430.
PR 14-OCT-1999; 99US-0159431.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161459.
PR 26-OCT-1999; 99US-0161460.
PR 26-OCT-1999; 99US-0161461.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2_COPY_178_186 x AAC40930 ..

Align seq 1/1 to: AAC40930 from: 1 to: 741

<GlyLysGluGlnCysAla 8

```
          ||||||||||||||||||
      433 GGAAAGGAACTTTGGGCT 450

Seq name: /SIDS1/gcgdata/gcgseq/genseqn/NA1992.DAT:AAQ27946

seq documentation block:
ID   AAQ27946 standard; cDNA; 787 BP.
XX
AC   AAQ27946;
XX
DT   28-JAN-1993  (first entry)
XX
DE   Sequence encoding P6 precursor protein.
XX
KW   Monocyte chemoattractant; bovine P6 derivative; thrombosis; tumour;
KW   inflammation therapy; ss.
XX
OS   Bos taurus.
XX
FH   Key             Location/Qualifiers
FT   misc_signal     42..48
FT                   /*tag = a
FT                   /label= translation consensus sequence
FT   polyA_signal    746..751
FT                   /*tag = b
FT   sig_peptide     45..113
FT                   /*tag = c
FT   mat_peptide     114..344
FT                   /*tag = d
XX
PN   DE4125251-C.
XX
PD   03-SEP-1992.
XX
PF   11-JUL-1991;   91DE-4125251.
XX
PR   11-JUL-1991;   91DE-4125251.
XX
PA   (SCHA-) SCHAPER & BRUEMMER GMBH & CO KG.
XX
PI   Gramm W,  Lins E;
XX
DR   WPI; 1992-293138/36.
DR   P-PSDB; AAR26581.
XX
PT   Drug containing a bovine protein homologous to human MCP-1 - for
PT   treating inflammation, tumours, thrombosis, and immune reactions,
PT   also for diagnosis
XX
PS   Claim 2; Page 4; 6pp; German.
XX
CC   Poly(A+)RNA from bull seminal vesicles was used to prepare a cDNA in
CC   the expression vector lambda gt11.  1.5 x 10(5) cDNA clones were
CC   screened with a polyclonal anti-P6 antiserum of monospecific
CC   immunoglobulin G and six positives were identified.  The insert of a
CC   suitable cDNA clone, pB42, was cloned into pUC18 and sequenced.
CC   pB42 encodes the 11,114 Da precursor of P6.  It is called Monocyte
CC   Chemoattractant (bMCP-1), which is a homologue of human (h)MCP-1.
CC   There is 72% overall AA sequence homology to hMCP-1 with the signal
CC   peptide showing 100% and the central region showing 89% homology.
XX
SQ   Sequence 787 BP; 236 A; 162 C; 144 G; 245 T; 0 other;


alignment scores:
            Quality:    6.00              Length:      6
              Ratio:   1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_178_186 x AAQ27946  ..

 Align seq 1/1  to: AAQ27946  from: 1  to: 787
```

```
          4 GlyLysGluLeuCysAla 8
            ||||||||||||||||||
      256 GGAAAGAGTTATGTGCA 272

Seq name: /...../Parental_Genseq/Genseq_NA1997.DAT:AAP95846

seq documentation block:
ID   AAP95846 standard; DNA; 999 BP.
XX
AC   AAP95846;
XX
DT   11-MAY-1998  (first entry)
XX
DE   Trichophaea saccata lipolytic enzyme DNA.
XX
KW   Lipolytic enzyme; lipase; alkalophilic enzyme; detergent;
KW   surfactant; esterification; lipid hydrolysis;
KW   optical isomer resolution; ss.
XX
OS   Trichophaea saccata CBS 804.70.
XX
FH   Key             Location/Qualifiers
FT   CDS             68..766
FT                   /*tag = a
FT   sig_peptide     68..160
FT                   /*tag = b
FT   mat_peptide     161..763
FT                   /*tag = c
FT                   /note= "Claim 22"
XX
PN   WO9741212-A1.
XX
PD   06-NOV-1997.
XX
PF   22-APR-1997;   97WO-DK00179.
XX
PR   25-APR-1996;   96DK-0000501.
PR   25-APR-1996;   96DK-0000500.
XX
PA   (NOVO-) NOVO NORDISK AS.
XX
PI   Borch K,  Balkier T,  Hirayama S,  Nielsen BR,  Oxenboll KM;
PI   Sandal T,  Taipar;
XX
DR   WPI; 1997-549717/50.
DR   P-PSDB; AAW38264.
XX
PT   Lipolytic enzymes useful as detergent additives at high pH - from
PT   Gliocladium, Verticillium and Trichophaea genera, give good washing
PT   performance and stability in detergent solutions
XX
PS   Claim 22; Page 43-45; 71pp; English.
XX
CC   This DNA sequence from Trichophaea saccata CBS 804.70 includes a
CC   coding region for a novel lipase (see AAW38264) that has lipolytic
CC   activity at alkaline pH and which gives good washing performance
CC   and stability in detergent solutions.  The lipase DNA was isolated
CC   by cloning a cDNA library from T. saccata in a vector, transforming
CC   yeast cells, and screening by expressed lipolytic activity for
CC   positive clones.  The isolated DNA can be used for recombinant
CC   production of the lipase, especially the mature enzyme.  The DNA is
CC   deposited in Escherichia coli transformant DSM 11298.  Claimed
CC   lipases (see also AAW38262 and AAW38263) can be used as detergent
CC   additives (e.g. as non-dusting granulates, stabilised liquids,
CC   slurries or protected enzymes) to remove lipid or fatty stains, and
CC   included with a surfactant in enzymatic detergent compositions
CC   (claimed) e.g. for laundry and dishwashing, especially at high pH.
CC   They are also useful for interesterification, total hydrolysis of
CC   fats and oils and optical isomer resolution.
XX
SQ   Sequence 999 BP; 260 A; 241 C; 239 G; 259 T; 0 other;
```

```
alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_178_186 x AAY95846.pro

 Align seq 1/1  to reverse of: AAY95846  from: 1  to: 999

        1 LeuAlaGlyLysGluLeu 6
          ||||||||||||||||||
      754 CTTGCTGGCAAGGAACTG 737

seq name: /tmp1/geneseq/geneseq-p-NA.fsa; TAI AAY95846

seq_documentation_block:
ID   AAZ32964 standard; DNA; 1620 BP.
XX
AC   AAZ32964;
XX
DT   09-FEB-2000  (first entry)
XX
DE   Human wild-type protein kinase MKK4 gene fragment B.
XX
KW   MKK4; mitogen activated protein kinase; MAPK; MAPK pathway; mutation;
KW   somatic; signal transduction; apoptosis; stress; cytokine; induction;
KW   phosphorylation; Jun kinase; JNK; p38; tumour; suppressor;
KW   loss of heterozygosity; LOH; cancer; detection; diagnosis; prognosis;
KW   breast cancer; pancreatic cancer; colorectal cancer; testicular cancer;
KW   drug screening; gene therapy; protein replacement therapy; mimetic; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   intron          1..1325
FT                   /*tag= a
FT   exon            1326..1428
FT                   /*tag= b
FT                   /note= "MKK4 exon B"
FT   intron          1429..1620
FT                   /*tag= c
XX
PN   US5989885-A.
XX
PD   23-NOV-1999.
XX
PF   13-JUN-1997;   97US-0874186.
XX
PR   10-JAN-1997;   97US-0782482.
XX
PA   (MYRI ) MYRIAD GENETICS INC.
XX
PI   Skolnick MH,  Perry WL,  Tavtigian SV.  Jeng DR ;
XX
DR   WPI; 2000-022786/02.
XX
PT   Polynucleotides comprising all or a portion of the tumor suppressor
PT   gene MKK4 locus are useful for diagnosis, prognosis and therapy of
PT   human cancers
XX
PS   Claim 1; Columns 63-66; 64pp; English.
XX
CC   This sequence represents a fragment of the gene encoding a human wild
CC   type protein kinase, MKK4, which includes exon B. The gene comprises 11
CC   coding exons. MKK4 (also known as SNPK1 and SEK1) may be involved in a
CC   MAPK (mitogen-activated protein kinase) pathway for the signal
CC   transduction of cytokine-induced and stress-induced apoptosis. MEK4 is
CC   also involved in suppressing a variety of tumours. MKK4 is a dual
CC   specific kinase that activates Jun kinases (JNKs) and p38 (a MAPK) but
CC   not extracellular signal-regulated kinases (ERKs) which are a subgroup
CC   of MAPKs. The JNK and p38 MAPKs are activated via dual phosphorylation
CC   on threonine and tyrosine and then go on to activate p: toins further
CC   downstream in signal transduction pathways. Tumour suppressor genes such
CC   as MKK4 are deleted at high frequency in certain tumour types. The
CC   deletions often involve loss of a single allele, which is known as loss
CC   of heterozygosity (LOH), and the remaining allele is presumed to be
CC   non-functional, either because of a pre-existing inherited mutation, or
CC   because of a secondary sporadic mutation. Alternatively, the deletion
CC   may involve homozygous deletion of both alleles. LOH events commonly
CC   involve deletions spanning many megabases of DNA, while homozygous
CC   deletions are relatively small in size, probably due to the proximity of
CC   essential genes. Sequences derived from the MKK4 gene can be used to
CC   detect a portion of the MKK4 locus or its expression product in a tissue
CC   sample for the diagnosis and prognosis of human cancer, and can
CC   also be used to diagnose of a predisposition to breast, pancreatic,
CC   colorectal and testicular cancers, as specific MKK4 mutations have
CC   been identified from such tumours. MKK4 oligonucleotides
CC   are useful for the detection of the nucleotide sequence of a particular
CC   MKK4 allele via PCR, and can be used as probes to detect point mutations,
CC   PCR amplification probes and mismatches between the MKK4 gene or mRNA.
CC   MKK4 protein can be used for screening of drugs which can restore MKK4
CC   gene product function for cancer therapy. MKK4 gene therapy, protein
CC   replacement therapy and protein mimetics that reconstitute the function
CC   of the MKK4 protein may be used for therapy of human cancers which result
CC   from a mutation in the MKK4 gene.
XX
SQ   Sequence 1620 BP; 427 A; 256 C; 273 G; 612 T; 2 other:


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_178_186 x AAZ32964  ..

 Align seq 1/1  to: AAZ32964  from: 1  to: 1620

        4 GlyLysGluLeuCysAla 8
          ||||||||||||||||||
      187 GGCAAGGAGCTCTGTGCT 204

seq name: /tmp1/geneseq/geneseq-p-NA.fsa; TAI AAX35660

seq_documentation_block:
ID   AAX35660 standard; DNA; 2063 BP
XX
AC   AAX35660;
XX
DT   16-JUL-1999  (first entry)
XX
DE   DNA encoding a potato acid invertase protein.
XX
KW   Localised mutation; target gene; plant cell;
KW   recombinogenic oligonucleobase; selective growth advantage;
KW   herbicide resistance; preservation; fruit; flower; potato; ss.
XX
OS   Solanum tuberosum.
XX
PN   WO9907865-A1.
XX
PD   18-FEB-1999.
XX
PF   05-AUG-1998;   98WO-US16267.
XX
PR   05-AUG-1997;   97US-0054836.
XX
PA   (KIME ) KIMERAGEN INC.
XX
PI   Arntzen CJ,  Kipp PB,  Kumar R,  May GD;
XX
DR   WPI; 1999-302251/25.
XX
```

PT Introducing mutations into target genes in plant cells using a
PT recombinogenic oligonucleobase comprising 2 regions homologous to a
PT target gene and an intervening mutant region
XX
PS Example 2; Page 36 47; 54pp; English.
XX
CC The specification describes methods for introducing localized mutations
CC into target genes in plant cells. The methods comprise using a
CC recombinogenic oligonucleobase comprising 2 regions homologous to a
CC target gene and an intervening mutant region. The methods can be used
CC to introduce localized mutations into target genes to introduce desirable
CC traits, e.g. selective growth advantage under appropriate selective
CC conditions, change in colour of plant cells growing in a callus,
CC herbicide resistance, or improved preservation of fruit or flowers.
CC They can be used in plants such as maize, wheat, rice, lettuce, potato,
CC tomato, canola, soybean or cotton cell. The present sequence represents
CC a target for the methods of the invention.
XX
SQ Sequence 2063 BP; 556 A; 458 C; 472 G; 577 T; 0 other;

Alignment scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

Alignment block:
US-09-126-945b-2_copy_178_186 x AAX35660 ..

Align seq 1/1 to: AAX35660 from: 1 to: 2063

1 LeuAlaGlyLysGluLeu 6
||||||||||||||||||
295 TTAGCTGGCAAAGAACTG 312

[illegible]

Date: Sep 14, 2001 9:54 PM

About: Results were produced by the [illegible]

Copyright (c) [illegible]

Command line parameters:
-MODEL frame+_p2n model -DEV xlp
-Q=/sw2/sw2_prod/SEQ_rpt_US/09126945/us-09-126-945b-2_copy_178_186.olip2n.rni /app/query/fasta_1.1267
-DB=Issued_Patents_NA -QFMT fastap -SUFFIX olip2n.rni
-GAPOP 4.500 -GAPEXT 0.050 -MINMATCH=0.100 -LOOPCL 0.000
-LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT 0.050 -XGAPOP 60.000
-XGAPEXT 60.000 -FGAPOP=6.000 -FGAPEXT 7.000 -YGAPOP 60.000
-YGAPEXT=60.000 -DELOP 6.000 -DELEXT=7.000 -START 1
-MATRIX oligo -TRANS=human40.cdi -LIST=45 -DOCALIGN 200
-THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE LOCAL -OUTFMT pts
-NORM=ext -MINLEN 0 -MAXLEN 2000000000
-USER US09126945_2COPY_1_91 -NCPU=6 -TCPU 3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-126-945B-2_COPY_178_186
Query length: 9
Database: Issued_Patents_NA:*
Database sequences: 324599
Database length: 94655562
Search time (sec): 322.720000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000

score_list:
Sequence Strd Orig ZScore EScore Len ! Documentation ..
[illegible]

seq_documentation block:
; Sequence 198, Application US/08991789A
; Patent No. 6225054
; GENERAL INFORMATION:
; APPLICANT: Friedakis, Tony N.
; Smith, John M.
; Reed, Steven G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF BREAST CANCER
; NUMBER OF SEQUENCES: 292
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed IP Law Group
; STREET: 701 Fifth Avenue, Suite 6300
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104 7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/991,789A
; FILING DATE: 11-Dec-1997
; CLASSIFICATION: Unknown
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E. R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 210121.419C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 198:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 299 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 198:
US-08-991-789A-198

alignment_scores:
quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

[illegible]

```
;     APPLICANT:  Tenq, David H.F.
;     APPLICANT:  Tavtigian, Sean V.
;     APPLICANT:  Perry III, William L.
;     APPLICANT:  Skolnick, Mark H.
;     TITLE OF INVENTION:  SPECIFIC MUTATIONS OF MAP KINASE KINASE
;     TITLE OF INVENTION:  4 (MKK4) IN HUMAN TUMOR CELL LINES IDENTIFY IT AS A TUMOR
;     TITLE OF INVENTION:  SUPPRESSOR IN VARIOUS TYPES OF CANCER
;     NUMBER OF SEQUENCES:  96
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Venable, Baetjer, Howard & Civiletti, LLP
;       STREET:  1201 New York Avenue, N.W., Suite 1000
;       CITY:  Washington
;       STATE:  DC
;       COUNTRY:  U.S.A.
;       ZIP:  20005
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/874,186
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/782,492
;       FILING DATE:  10-JAN-1997
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Saxe, Stephen A.
;       REGISTRATION NUMBER:  38,609
;       REFERENCE/DOCKET NUMBER:  24884-121492-01
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  202-962-4848
;       TELEFAX:  202-962-8300
;   INFORMATION FOR SEQ ID NO:  37:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1620 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     FEATURE:
;       NAME/KEY:  intron
;       LOCATION:  1..1425
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  1326..1428
;     FEATURE:
;       NAME/KEY:  intron
;       LOCATION:  1429..1620
US-08-874,186-37


alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_178_186 x US-08-874,186-37  .

 Align seq 1/1  to reverse of: US-08-874,186-37  from: 1  to: 1520

        3 GlyLysGluLeuCysAla 8
          ||||||||||||||||||
      187 GGCAAGGAGCTCTGTGCT 204

seq name: /projects/data/Data_Submission/US_09MB/seq_rev_2.seq  rev 2-Jun-2001 ;

seq documentation block:
; Sequence 4, Application US/08866650

; Patent No. 5939421
;  GENERAL INFORMATION:
;     APPLICANT:  Greenspan, Daniel S
;     APPLICANT:  Takahara, Kazuhiko
;     APPLICANT:  Bottman, Guy G
;     TITLE OF INVENTION:  Mammalian Tolloid Like Protein
;     NUMBER OF SEQUENCES:  13
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Quarles & Brady
;       STREET:  1 South Pinckney Street
;       CITY:  Madison
;       STATE:  WI
;       COUNTRY:  US
;       ZIP:  53703
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/866,650
;       FILING DATE:
;       CLASSIFICATION:  514
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Berson, Bennett J
;       REGISTRATION NUMBER:  37094
;       REFERENCE/DOCKET NUMBER:  960296.93839
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  608 251 5000
;       TELEFAX:  608 251 9166
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3919 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     ORIGINAL SOURCE:
;       ORGANISM:  Homo sapiens
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  648..3689
;       OTHER INFORMATION:  /product= "human mTll protein"
US-08-866,650-4


alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_178_186 x US-08-866,650-4/rev  .

 Align seq 1/1  to reverse of: US-08-866,650-4  from: 1  to: 3919

        2 AlaGlyLysGluLeuCys 7
          ||||||||||||||||||
      405 GCAGGTAAGGAACTCTGC 388

seq name: /projects/data/Data_Submission/US_09MB/seq_rev_2.seq;US-09-021,287-4

seq documentation block:
; Sequence 4, Application US/09021287
; Patent No. 5981717
;  GENERAL INFORMATION:
;     APPLICANT:  Greenspan, Daniel S
;     APPLICANT:  Takahara, Kazuhiko
;     APPLICANT:  Bottman, Guy G
;     TITLE OF INVENTION:  Mammalian Tolloid-like Protein
;     NUMBER OF SEQUENCES:  13
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Quarles & Brady
```

```
;       STREET:  1 South Pinckney Street
;       CITY:  Madison
;       STATE:  WI
;       COUNTRY:  US
;       ZIP:  53703
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/021,287
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/866,650
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Berson, Bennett J
;       REGISTRATION NUMBER:  37094
;       REFERENCE/DOCKET NUMBER:  960296.94865
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  608-251-5000
;       TELEFAX:  608-251-9166
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4919 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     ORIGINAL SOURCE:
;       ORGANISM:  Homo sapiens
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  648..3689
;       OTHER INFORMATION:  /product= "human mill protein"
US-09-021-287-4


alignment_scores:
              Quality:     6.00              Length:       6
                Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_178_186 x US-09-021-287-4/rev   ..

 Align seq 3/1  is reverse of  US-09-021-287-4  from  1  to  3414

        2 AlaGlyLysGluLeuCys 7
          ||||||||||||||||||
      405 GCAGGTAAGGAACTCTGC 388

Seq_name:  [illegible]

seq_documentation_block:
; Sequence 59, Application US/08174672D
; Patent No. 5877009
;  GENERAL INFORMATION:
;     APPLICANT:  Zannis Ph.D., Vassilis I.
;     APPLICANT:  Cladaras Ph.D., Christos
;     TITLE OF INVENTION:  APOLIPOPROTEIN GENE REGULATION
;     NUMBER OF SEQUENCES:  113
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Choate, Hall & Stewart
;       STREET:  53 State Street
;       CITY:  Boston
;       STATE:  MA
;       COUNTRY:  USA
;       ZIP:  02109-2891
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/174,672D
;       FILING DATE:  28 DEC 1994
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Jartoli Ph.D., Brenda H.
;       REGISTRATION NUMBER:  39,223
;       REFERENCE/DOCKET NUMBER:  0079571 0005
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 248 5000
;       TELEFAX:  (617) 248 4000
;   INFORMATION FOR SEQ ID NO:  59:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  26 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  both
;       TOPOLOGY:  not relevant
;     MOLECULE TYPE:  DNA (genomic)
;     IMMEDIATE SOURCE:
;       CLONE:  AIII oligonucleotide
US-08-174-672D-59


alignment_scores:
              Quality:     5.00              Length:       5
                Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_178_186 x US-08-174-672D-59/rev   ..

 Align seq 2/1  is reverse of  US-08-174-672D-59  from  1  to  26

        1 LeuAlaGlyLysGlu 5
          ||||.||||||||||
       21 CTGGCAGGAAAAGAG 7

Seq_name:  [illegible]

seq_documentation_block:
; Sequence 104, Application: US/09425416A
; Patent No. 6210942
; GENERAL INFORMATION:
;  APPLICANT:  Lewis, No. 6210942man G.
;  APPLICANT:  Davin, Laurence B.
;  APPLICANT:  Stevenel-Kostova, Albena T.
;  APPLICANT:  Fujita, Masayuki
;  APPLICANT:  Gang, David R.
;  APPLICANT:  Sarkanen, Simo
;  APPLICANT:  Ford, Joshua D
;  TITLE OF INVENTION:  RECOMBINANT PINORESINOL/LARICIRESINOL REDUCTASES,
;  TITLE OF INVEN. N:  RECOMBINANT DIRECTED REFERENCE AND METHODS OF USE
;  FILE REFERENCE:  WSUR 1-13793
;  CURRENT APPLICATION NUMBER:  US/09/475,416A
;  CURRENT FILING DATE:  1999-12-30
;  PRIOR APPLICATION NUMBER:  09/307,653
;  PRIOR FILING DATE:  1999-05-07
;  PRIOR APPLICATION NUMBER:  PCT/US97/20491
;  PRIOR FILING DATE:  1997-11-07
;  PRIOR APPLICATION NUMBER:  60/054,380
;  PRIOR FILING DATE:  1997-07-31
;  PRIOR APPLICATION NUMBER:  60/030,522
;  PRIOR FILING DATE:  1996-11-08
;  NUMBER OF SEQ ID NOS:  122
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 104
;   LENGTH: 29
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
```

```
;    OTHER INFORMATION: Description of Artificial Sequence:
;    OTHER INFORMATION: oligonucleotide
;    NAME/KEY: misc_feature
;    LOCATION: (1)..(29)
;    OTHER INFORMATION: primer SpN2
US-09-475-416A-104


alignment scores:
              Quality:     5.00              Length:       5
                Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2_COPY_178_186 x US-09-475-416A-104/rev   ..

 Align seq 1/1   to reverse of: US-09-475-416A-104   from: 1   to: 29

        4 GlyLysGluLeuCys 7
          |||||||||||||||
       26 GCAAAAGAATTATGT 12

Seq name: /cgn2_1/pt_data/_tmp_6A_.comMs.seqpop.97.600.rls.175

Seq documentation block:
; Sequence 175, Application US/09009913
; Patent No. 6087485
;  GENERAL INFORMATION:
;     APPLICANT:  Axys Pharmaceuticals, Inc.
;     TITLE OF INVENTION:  Asthma Related Genes
;     NUMBER OF SEQUENCES:  339
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Bozicevic & Reed, LLP
;       STREET:  285 Hamilton Ave, Suite 200
;       CITY:  Palo Alto
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  94301
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/009,913
;       FILING DATE:  21 JAN 1998
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Sherwood, Pamela J
;       REGISTRATION NUMBER:  36,677
;       REFERENCE/DOCKET NUMBER:  SEQ-4P
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  650-327-3231
;       TELEFAX:  650 327 3231
;       TELEX:
;  INFORMATION FOR SEQ ID NO:  175:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH:  30 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-09-009-913-175


alignment scores:
              Quality:     5.00              Length:       5
                Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2_COPY_178_186 x US-09-009-913-175   ..

 Align seq 1/1   to: US-09-009-913-175   from: 1   to: 30

        1 LeuAlaGlyLysGlu 5
          ||| ||||||||||
        7 TTGGCTGGTAAGGAA 21

Seq name: /cgn2_1/pt_data/_tmp_6A_.comMs.seqpop.97.600.rls.175

Seq documentation block:
; Sequence 30, Application US/08132990A
; Patent No. 5834589
;  GENERAL INFORMATION:
;     APPLICANT:  MERMELOD, DANIEL
;     APPLICANT:  YOSHIMOTO, TAKAYUKI
;     TITLE OF INVENTION:  Human Retrovirus Receptor and DNA Coding Therefor
;     NUMBER OF SEQUENCES:  31
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Pennie & Edmonds
;       STREET:  1155 Avenue of the Americas
;       CITY:  New York
;       STATE:  NY
;       COUNTRY:  USA
;       ZIP:  10036
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  Patent In Release #1.24
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/132,990A
;       FILING DATE:  07-OCT-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/084,729
;       FILING DATE:  29-JUN-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US93/05569
;       FILING DATE:  11-JUN-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/899,075
;       FILING DATE:  11-JUN-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/806,178
;       FILING DATE:  13-DEC-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/627,960
;       FILING DATE:  14-DEC-1990
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Mistrock, S. Leslie
;       REGISTRATION NUMBER:  18,872
;       REFERENCE/DOCKET NUMBER:  8105.004.999
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 790-9090
;       TELEFAX:  (212) 869-8864
;       TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  30:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  38 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA
US-08-132-990A-30


alignment scores:
              Quality:     5.00              Length:       5
                Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2_COPY_178_186 x US-08-132-990A-30   ..
```

```
 Align seq 1/1  to: US-08-132-990A-30  from: 1  to: 38

       2 AlaGlyLysGluLeu 6
         ||||||||||||||||
      17 GCTGGGAAGGAGCTA 31

Seq_name: /tmp2_%j*_dir/_tra_%B_*.MS_seq*US-08-132-990A-31

seq_documentation_block:
; Sequence 31, Application US/08132990A
; Patent No. 5834589
;  GENERAL INFORMATION:
;     APPLICANT:  MERUELO, DANIEL
;     APPLICANT:  YOSHIMOTO, TAKAYUKI
;     TITLE OF INVENTION:  Human Retrovirus Polymerase cDNA Clones
;     NUMBER OF SEQUENCES:  31
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Pennie & Edmonds
;       STREET:  1155 Avenue of the Americas
;       CITY:  New York
;       STATE:  NY
;       COUNTRY:  USA
;       ZIP:  10036
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  Patent In Release #1.24
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/132,990A
;       FILING DATE:  07-OCT-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/084,729
;       FILING DATE:  29-JUN-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US93/05569
;       FILING DATE:  11-JUN-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/899,875
;       FILING DATE:  11-JUN-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/806,178
;       FILING DATE:  13-DEC-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/627,959
;       FILING DATE:  14-DEC-1990
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Misrock, S. Leslie
;       REGISTRATION NUMBER:  18,872
;       REFERENCE/DOCKET NUMBER:  8105-004-999
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 790-9090
;       TELEFAX:  (212) 869-8864
;       TELEX:  66141 PENNIE
;   INFORMATION FOR SEQ ID NO:  31:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  38 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA
US-08-132-990A-31


alignment_scores:
            Quality:     5.00              Length:       5
              Ratio:  1.000                Gaps:         0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_178_186 x US-08-132-990A-31    .

 Align seq 1/1  to: US-08-132-990A-31  from: 1  to: 38

       2 AlaGlyLysGluLeu 6
         ||||||||||||||||
      17 GCTGGGAAGGAGCTA 31

Seq_name: /tmp_%j*_dir/_tra_%B_*.MS_seq*US-08-132-990A-31

seq_documentation_block:
;Patent No. 5256648
;     APPLICANT: MARSAH, FRANCIS P.;EDELSON,RICHARD L.
;     TITLE OF INVENTION: SELECTIVE INHIBITION OF GENE EXPRESSION
;BY PHOTOACTIVATABLE OLIGONUCLEOTIDES
;     NUMBER OF SEQUENCES: 40
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/410,622
;       FILING DATE: 21-SEP-1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 299,265
;       FILING DATE: 23-JAN-1989
;       APPLICATION NUMBER: 146,571
;       FILING DATE: 21-JAN-1988
;SEQ ID NO:34:
;       LENGTH:  39
5256648-34


alignment_scores:
            Quality:     5.00              Length:       5
              Ratio:  1.000                Gaps:         0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_178_186 x 5256648-34    ..

 Align seq 1/1  to: 5256648-34  from: 1  to: 39

       3 GlyLysGluLeuCys 7
         |||||||||||||||
       5 GGGAAAGAACTGTGT 19

Seq_name: /tmp_%j*_dir/_tra_%B_*.MS_seq*US-09-332-934-4

seq_documentation_block:
; Sequence 4, Application US/09332934
; Patent No. 6114123
; GENERAL INFORMATION:
;  APPLICANT: Murry, Lynn, E.
;  APPLICANT: Tang, Tom, Y.
;  APPLICANT: Baughn, Marian, R.
;  TITLE OF INVENTION: LIPOCALIN FAMILY PROTEIN
;  FILE REFERENCE: PC 0005 US
;  CURRENT APPLICATION NUMBER: US/09/332,934
;  CURRENT FILING DATE: 1999-06-14
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PERL Program
; SEQ ID NO 4
;   LENGTH: 260
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY:
;   OTHER INFORMATION: 441944381
;  PUBLICATION INFORMATION:
US-09-332-934-4


alignment_scores:
            Quality:     5.00              Length:       5
              Ratio:  1.000                Gaps:         0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
```

```
 US 09 126 945B 2 COPY_178_186 x US 09 332 934 4.rny   ..

 Align seq 1/1  to reverse of: US 09 332 934 4  from: 1  to: 260

        2 AlaGlyLysGluLeu 6
          ||||||||||||||||
      184 GCTGGAAAAGAGCTC 169

Seq name: /pn1/pt1 dir2/rice/US 09 MB.seq.gz.rny.gz 174 6725 1

seq documentation block:
; Sequence 1, Application US/08174672D
; Patent No. 5877009
;  GENERAL INFORMATION:
;     APPLICANT:  Zannis Ph.D., Vassilis I.
;     APPLICANT:  Cladaras Ph.D., Christos
;     TITLE OF INVENTION:  APOLIPOPROTEIN GENE REGULATION
;     NUMBER OF SEQUENCES:  113
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Choate, Hall & Stewart
;       STREET:  53 State Street
;       CITY:  Boston
;       STATE:  MA
;       COUNTRY:  USA
;       ZIP:  02109 2804
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/174,672D
;       FILING DATE:  28 DEC 1993
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Jarrell Ph.D., Brenda H.
;       REGISTRATION NUMBER:  39,223
;       REFERENCE/DOCKET NUMBER:  0029571 0005
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 248-5000
;       TELEFAX:  (617) 248 4000
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  262 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  both
;       TOPOLOGY:  not relevant
;     MOLECULE TYPE:  DNA (genomic)
;     IMMEDIATE SOURCE:
;       CLONE:  apoA II promoter region  911 to  650
US 08 174 672D 1


alignment scores:
              Quality:     5.00              Length:       5
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US 09 126 945B 2_COPY_178_186 x US 08 174 672D 1.rny   ..

 Align seq 1/1  to reverse of: US 08 174 672D 1  from: 1  to: 262

        1 LeuAlaGlyLysGlu 5
          |||||||||||||||
      226 CTGGCAGGAAAACAC 212

Seq name: /pn1/pt1 dir2/rice/PCT MB.seq.gz.rny.gz 05/5 16766 4

seq documentation block:
; Sequence 4, Application PC/TUS9516766
;  GENERAL INFORMATION:
;     APPLICANT:  COLD SPRING HARBOR, Laboratories
;     TITLE OF INVENTION:  CANCER DETECTION PROBES
;     NUMBER OF SEQUENCES:  35
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;       STREET:  Four Embarcadero Center, Suite 3400
;       CITY:  San Francisco
;       STATE:  CA
;       COUNTRY:  US
;       ZIP:  94111
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/16766
;       FILING DATE:  12 DEC 1995
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  ROWLAND, Bertram I
;       REGISTRATION NUMBER:  20015
;       REFERENCE/DOCKET NUMBER:  FP60430 I PC
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415 781 1989
;       TELEFAX:  415 398 3249
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  450 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  other nucleic acid
;       DESCRIPTION:  /desc = "oligonucleotide"
PCT US95 16766 4


alignment scores:
              Quality:     5.00              Length:       5
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US 09 126 945B 2_COPY_178_186 x PCT US95 16766 4.rny   ..

 Align seq 1/1  to reverse of: PCT-US95-16766-4  from: 1  to: 350

        1 LeuAlaGlyLysGlu 5
          |||| ||||||||||
       47 TTAGCTGGAAAGGAG 33

Seq name: /pn1/pt1 dir2/rice/US MB.seq.gz.rny.gz 277 565 15

seq documentation_block:
; Sequence 15, Application US/09277565
; Patent No. 6207384
; GENERAL INFORMATION:
;  APPLICANT: Mekalanos, John J.
;  APPLICANT: Akerley, Brian J.
;  APPLICANT: Rubin, Eric J.
;  APPLICANT: Camilli, Andrew
;  TITLE OF INVENTION: SYSTEMATIC IDENTIFICATION OF ESSENTIAL
;  TITLE OF INVENTION: GENES BY IN VITRO TRANSPOSON MUTAGENESIS
;  FILE REFERENCE: 00742/052002
;  CURRENT APPLICATION NUMBER: US/09/277,565
;  CURRENT FILING DATE: 1999-03-26
;  EARLIER APPLICATION NUMBER: 60/079,270
;  EARLIER FILING DATE: 1998-03-27
;  NUMBER OF SEQ IDS NOS: 29
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
;   LENGTH: 481
;   TYPE: DNA
;   ORGANISM: Haemophilus influenzae
```

US-09-277-565-15

alignment_scores:
Quality: 5.00 Length: 5
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_178_186 x US-09-277-565-15 ..

Align seq 1/1 to: US-09-277-565-15 from: 1 to: 381

2 AlaGlyLysGluLeu 6
|||||||||||||||
3 GCAGGAAAAGAACTT 17

Seq_name: ;`ip2_f;p'=d1*a/2;'i=,'A_COMP seq:US-08-594-031-76

seq_documentation_block:
; Sequence 76, Application US/08594031
; Patent No. 5783182
; GENERAL INFORMATION:
; APPLICANT: THOMPSON, Timothy C.
; TITLE OF INVENTION: METHOD FOR IDENTIFYING METASTATIC SEQUENCES
; NUMBER OF SEQUENCES: 175
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BAKER & BOTTS, L.L.P.
; STREET: 1299 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004-2400
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/594,031
; FILING DATE: 30-JAN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/006,838
; FILING DATE: 16-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Remenick, James
; REGISTRATION NUMBER: 36,902
; REFERENCE/DOCKET NUMBER: 0A146-0110
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-639-7700
; TELEFAX: 202-639-7890
; TELEX:
; INFORMATION FOR SEQ ID NO: 76:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 383 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
US-08-594-031-76

alignment_scores:
Quality: 5.00 Length: 5
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-126-945B-2_COPY_178_186 x US-08-594-031-76/rev ..

Align seq 1/1 to reverse of: US-08-594-031-76 from: 1 to: 383

1 LeuA.aGlyLysGlu 5
||.| ||||||||||
270 TTGGCTGGAAAAGAA 256

```
CMD of: US-09-126-945B-2_COPY_178_186 to: EST:*  -out_format : pts

Date: Sep 14, 2001  7:47 PM

About: Results were produced by the GenScore software, version 4.5.
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
  -MODEL frame+_p2n.model -DEV xlp
  -Q=/cgn2/tmp/US09_sptb/US09_6645/runat_15%1_1/185271_1645/app_proxy_file_1_1267
  -DB EST -FORMT=tastap -SUFFIX olip2n.rst -GAPOP=4.500
  -GAPEXT 0.050 -MINMATCH=6.000 -LOWPCT 0.000 -LOWMOST 0.000
  -QGAPOP=4.500 -QGAPEXT 0.050 -XGAPOP=60.000 -XGAPEXT 60.000
  -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT 60.000
  -DELOP 6.000 -DELEXT=7.000 -START 1 -MATRIX oligo
  -TRANS human40.cdi -LIST=45 -DoCALIGN 200 -THR_SCORE quality
  -THR_MIN=1 -ALIGN=15  MODE LOCAL -OUTFMT pts -NORM ext  MINLEN 0
  -MAXLEN=2000000000 -USER=US09126945B_%%IN1_1_4641 -NCPU 6
  -ICPU 3 -LONGLOG -NO_XLPXY -WAIT -THREADS 1

Search information block:
Query: US-09-126-945B-2_COPY_178_186
Query length: 9
Database: EST:*
Database sequences: 10228115
Database length: 431459454
Search time (sec): 9087.630000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming: YGAPEXT XGAPEXT 60.000


score_list:
Sequence                Strd Orig    ZScore    EScore Len  ! Documentation ..
qb_est97:BG179568         -    9.00  176.12   1.08      991  ! BG179568 602428182F1 NIH_MGC_91
qb_qss20:AZ065549         -    7.00  141.22   61.86     293  ! AZ065549 RPCI-23 431G15.TJ RPCI
qb_qss45:B50881           +    7.00  142.94   76.47     357  ! B50881 C11978SK-15304.TF C11978
qb_qss6:AQ062808          +    7.00  142.28   83.20     395  ! AQ062808 CIT-HSP-2349e16.TR CIT
qb_qss16:AQ804599         -    7.00  142.20   84.07     400  ! AQ804599 HS_2263_B1_B05_1A1 CIT
qb_est2:AA139849          +    7.00  141.65   90.16     435  ! AA139849 mq91a06.r1 Stratagene
qb_est23:A1671776         +    7.00  141.36   93.50     455  ! A1671776 wa05b01.x1 NCI_CGAP_K
qb_qss2:AZ256508          -    7.00  141.18   95.83     468  ! AZ256508 RPCI-23 222N21.TJ RPCI
qb_est78:BE757615         +    7.00  140.26  107.80     539  ! BE757615 212104 MARC 2B0V Bos t
qb_qss28:AZ560668         -    7.00  139.74  115.25     584  ! AZ560668 RPCI-23-219I12.TV RPCI
qb_est84:BF156352         +    7.00  139.16  124.06     638  ! BF156352 t151a04.y1 Sugano Kawa
qb_qss27:AZ470308         -    7.00  139.01  126.49     653  ! AZ470308 IMR0284-013F Mouse 10Kb
qb_est85:BF243620         +    7.00  137.28  158.03     953  ! BF243620 601879169F1 NIH_MGC_59
qb_est89:BF579116         +    7.00  131.53  330.04    2064  ! BF579116 602096050F1 NCI_CGAP_
qb_est18:AI418348         +    6.00  130.92  357.22     106  ! AI418348 ta78e01.x1 NCI_CGAP_HS
qb_est84:BF172211         +    6.00  128.28  500.89     159  ! BF172211 PCL4410 Myeloma (PCL)
qb_est42:AW124587         +    6.00  128.16  508.69     162  ! AW124587 UI-M-BH2-lsapt-t-05-0
qb_qss27:AZ473289         -    6.00  127.93  521.51     168  ! AZ473289 IMR0280A08F Mouse 10Kb
qb_est54:AW999360         -    6.00  126.70  613.92     203  ! AW999360 MR0-BN0070-230308-006
qb_est18:AI418595         +    6.00  126.54  626.49     208  ! AI418595 tb02d04.x1 NCI_CGAP_Lu
qb_est19:AI432766         +    6.00  126.23  651.50     218  ! AI432766 qp92h08.x1 Soares_fetu
qb_est54:BB199029         +    6.00  126.20  653.99     219  ! BB199029 BB199029 RIKEN full-l
qb_qss27:AZ495729         -    6.00  126.11  661.41     222  ! AZ495729 IMR0331-013K Mouse 10Kb
qb_est10:AA692296         +    6.00  126.08  663.92     223  ! AA692296 ot20d09.r1 Barstead ma
qb_est89:BF560056         +    6.00  126.06  666.41     224  ! BF560056 UI-R-A1-ev-d-11-0-UI.r
qb_est6:AA472909          -    6.00  125.97  673.83     227  ! AA472909 EST84878 Colon adenoca
qb_est55:BB049889         -    6.00  125.83  686.18     232  ! BB049889 BB049889 RIKEN full-l
qb_est95:BG019764         -    6.00  125.83  686.18     232  ! BG019764 de44c04.x1 NICHD_XGC_
qb_est33:AV162941         -    6.00  125.74  693.57     235  ! AV162941 AV162941 Mus musculus
qb_est61:BB299917         -    6.00  125.69  698.18     237  ! BB299917 BB299917 RIKEN full-l
qb_est20:A1415358         -    6.00  125.66  700.94     238  ! A1415358 md56b04.x1 Soares_mous
qb_est94:BB006775         -    6.00  125.66  700.94     238  ! BB006775 BB006775 RIKEN full-l
qb_est63:BB383491         +    6.00  125.66  700.94     238  ! BB383491 BB383491 RIKEN full-l
qb_est81:BF939069         +    6.00  125.63  703.49     239  ! BF939069 PM0-TN0121-260800-001
qb_est44:AV187389         +    6.00  125.61  705.84     240  ! AV187389 AV187389 Yuji Kohara
qb_est45:AV251012         -    6.00  125.61  705.84     240  ! AV251012 AV251012 RIKEN full-l
qb_est103:C44985          +    6.00  125.61  705.84     240  ! C44985 C44985 Yuji Kohara unpu
qb_est48:AV371243         -    6.00  125.55  710.74     242  ! AV371243 AV371243 RIKEN full-l
qb_est47:AV323074         -    6.00  125.53  713.19     243  ! AV323074 AV323074 RIKEN full-l
qb_est101:BG512115        -    6.00  125.53  713.19     243  ! BG512115 da42e06.y1 Weltcom
qb_est13:AA885066         -    6.00  125.50  715.64     244  ! AA885066 am11d04.s1 Soares_N
qb_est111:C12507          -    6.00  125.47  718.08     245  ! C12507 BPCOM5011 normalized
qb_qss7:AQ151144          -    6.00  125.37  727.84     249  ! AQ151144 HS_2186_A2_B11_MF_C
qb_est11:AA717293         -    6.00  125.32  732.70     251  ! AA717293 ot15h07.r1 Barstead
qb_est45:AW336266         -    6.00  125.14  749.69     258  ! AW336266 23-G6 MAFC 2B0V Bos

seq_name: qb_est97:BG179568

seq_documentation_block:
LOCUS       BG179568          991 bp    mRNA            EST       06 FEB 2001
DEFINITION  602428182F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4429413 5',
            mRNA sequence.
ACCESSION   BG179568
VERSION     BG179568.1  GI:12686271
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 991)
  AUTHORS   NIH MGC Project.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL)
            cDNA Sequenced by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: TLAM10181  row: d  column: 22
            High quality sequence stop: 655.
FEATURES             Location/Qualifiers
     source          1..991
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4429413"
                     /clone_lib="NIH_MGC_91"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: prostate; Vector: pCMV-SPORT6; Site_1: Not1;
                     Site_2: Sall; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.4 kb. Library was hed for
                     full length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      190 a    451 c    305 g    145 t
ORIGIN


alignment_scores:
            Quality:    9.00              Length:      9
              Ratio:   1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_178_186 x BG179568    ..

 Align seq 1/1   to: BG179568  from: 1  to: 991

       1 IenAlaGlyLysGluIenCysAlaMet 9
         ||||||||||||||||||||||||||
     435 CTGGCGGGCAAGGAGCTGTGCGCCATG 461

seq_name: qb_qss20:AZ065549

seq_documentation_block:
LOCUS       AZ065549          293 bp    DNA             GSS       30 MAR 2000
DEFINITION  RPCI-23 431G15.TJ RPCI-23 Mus musculus genomic clone RPCI-23-431G15
            , DNA sequence.
ACCESSION   AZ065549
```

```
VERSION     AZ065549.1  GI:7456861
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 293)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Lewins,M., Mequan,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI 23
  JOURNAL   Unpublished (1999)
COMMENT     other_GSSs: RPCI 23 431G15.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI 23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/bacinfo.htm)
            or from Research Genetics (info@resgen.com). BAC end paper:
            http://www.tigr.org/tdb/mgc/bac_ends/rpci23.html
            Plate: 431  row: G  column: 15
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..293
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23 431G15"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site 1:
                     EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and partially digested
                     with a combination of EcoRI and EcoRI Methylase.  Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI sites.  The ligation products were transformed into
                     DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT       82 a      69 c      60 g      82 t
ORIGIN


alignment_scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_178_186 x AZ065549/rev

 Align seq 1/1  to reverse of: AZ065549  from: 1  to: 293

        1 LeuAlaGlyLysGluLeuCys 7
          |||||||||||||||||||||
      293 CTTGCGGGTAAACAGCTCTGC 273

seq name: qb_qss45:B50881

seq documentation_block:
LOCUS       B50881          457 bp    DNA              GSS       18-JUN-1998
DEFINITION  CIT978SK 15403,TF CIT978SK Homo sapiens genomic clone 15403, DNA
            sequence.
ACCESSION   B50881
VERSION     B50881.1  GI:2603118
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 457)
  AUTHORS   Kim,U.-J., Adams,M.D. and Simon,M.I.
  TITLE     Determination of clone end sequences of human Bacterial Artificial
            Chromosomes
  JOURNAL   Unpublished (1997)
COMMENT     other_GSSs: CIT978SK 15403.TV
            Contact: Ung-Jin Kim
            Caltech Genome Research Lab
            California Institute of Technology
            Division of Biology, MS 147-75, Pasadena, CA 91125, USA
            Tel: 626-796-7066
            Fax: 626-395-4901
            Email: ung@ash.tree.caltech.edu
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/human_bac_end_seq/bac_end_search.html
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..457
                     /organism="Homo sapiens"
                     /db_xref="GDB:5280030"
                     /db_xref="taxon:9606"
                     /clone="15403"
                     /clone_lib="CIT978SK"
                     /sex="Female"
                     /cell_type="Fibroblast"
                     /note="Vector: pBAC108L; Site_1: HindIII; Site_2: HindIII;
                     CalTech Human BAC Library A"
BASE COUNT       65 a      74 c      68 g     150 t
ORIGIN


alignment_scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_178_186 x B50881  ..

 Align seq 1/1  to: B50881  from: 1  to: 457

        1 LeuAlaGlyLysGluLeuCys 7
          |||| ||||||||||||||||
      201 CTTGCTGGCAAGGAGCTGTGT 221

seq name: qb_qss6:AQ06280B

seq documentation_block:
LOCUS       AQ06280B        395 bp    DNA              GSS       31-JUL-1998
DEFINITION  CIT-HSP-2349O16.TR CIT-HSP Homo sapiens genomic clone 2349O16, DNA
            sequence.
ACCESSION   AQ06280B
VERSION     AQ06280B.1  GI:4364720
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 395)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Lieber,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shiroya,B.,
            Simon,M. and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence Ready Map
            Building (1998)
  JOURNAL   Unpublished (1998)
COMMENT     other_GSSs: CIT-HSP-2349O16.TF
            Contact: Mark Adams
            Department of Eukaryotic Genomics
```

```
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_ends/bac_ends.html
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..395
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone="2349016"
                     /clone_lib "CIT-HSP"
                     /sex "Male"
                     /cell_type="Sperm"
                     /note "Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
BASE COUNT       72 a     83 c     82 g    158 t
ORIGIN


alignment_scores:
            Quality:    7.00             Length:       7
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-9458-2_copy_178_186 x AQ062808  ..

 Align seq 1/1  to: AQ062808  from: 1  to: 395

       1 LeuAlaGlyLysGluLeuCys 7
         |||||||||||||||||||||
     174 CTTGCTGGTAAGGAGTTATGT 194

seq_name: qb_gss16:AQ804599

seq_documentation_block:
LOCUS        AQ804599       400 bp    DNA             GSS       09-AUG-1999
DEFINITION   HS_2263_B1_B05_T7C CIT Approved Human Genomic Sperm Library D Homo
             sapiens genomic clone Plate 2263 Col 9 Row D, DNA sequence.
ACCESSION    AQ804599
VERSION      AQ804599.1  GI:5721931
KEYWORDS     GSS.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 400)
  AUTHORS    Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
             Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
             Hood,L.
  TITLE      Sequence-tagged connectors: A sequence approach to mapping and
             scanning the human genome
  JOURNAL    Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE    99380589
COMMENT      Contact: Mahairas GG, Wallace JC, Hood L
             High Throughput Sequencing Center
             University of Washington
             401 Queon Anne Avenue North, Seattle, WA 98109, USA
             Tel: (206) 616-3618
             Fax: (206) 616-3887
             Email: jwallace@u.washington.edu
             Clones may be purchased from Research Genetics (info@resgen.com).
             BAC end Web Server: http://www.htsc.washington.edu
             Plate: 2263  row: D  column: 9
             Seq primer: T7
             Class: BAC ends
             High quality sequence stop: 400.
FEATURES             Location/Qualifiers
     source          1..400
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone "Plate 2263 Col 9 Row D"
                     /clone_lib "CIT Approved Human Genomic Sperm Library D"
                     /sex "male"
                     /note "Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E Coli DH10B"
BASE COUNT       145 a     95 c     74 g     86 t
ORIGIN


alignment_scores:
            Quality:    7.00             Length:       7
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-9458-2_COPY_178_186 x AQ804599/rev  ..

 Align seq 1/1  to reverse of: AQ804599  from: 1  to: 400

       1 LeuAlaGlyLysGluLeuCys 7
         |||.|||||||||||||||||
     309 CTTGCTGGTAAGGAGTTATGT 289

seq_name: qb_est2:AA139849

seq_documentation_block:
LOCUS        AA139849       435 bp    mRNA            EST       11-FEB-1997
DEFINITION   mq91a06.r1 Stratagene mouse heart (#937316) Mus musculus cDNA clone
             IMAGE:586066 5', mRNA sequence.
ACCESSION    AA139849
VERSION      AA139849.1  GI:1702053
KEYWORDS     EST.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 435)
  AUTHORS    Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
             Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
             Schellenberg,K., Stoptoe,M., Tan,E., Underwood,K., Moore,P.,
             Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
             Waterston,R.
  TITLE      The Wash-HHMI Mouse EST Project
  JOURNAL    Unpublished (1996)
COMMENT      Contact: Marra M/Mouse EST Project
             Wash-HHMI Mouse EST Project
             Washington University School of Medicine
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: mouseest@watson.wustl.edu
             This clone is available royalty free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             MGI:360714
             Seq primer: -28m13 rev1 ET from Amersham
             High quality sequence stop: 405.
FEATURES             Location/Qualifiers
     source          1..435
                     /organism "Mus musculus"
                     /strain: "NIH/Swiss"
                     /db_xref "taxon:10090"
                     /clone="IMAGE:586066"
                     /clone_lib="Stratagene mouse heart (#937316)"
                     /sex "pooled"
                     /tissue_type "heart"
                     /dev_stage "13 day embryos"
                     /lab_host "SOLR (kanamycin resistant)"
                     /note "Organ: heart; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     -21m13; dT, 93 pooled NIH/Swiss 13 day embryo hearts.
```

Average insert size: 1.0 kb; Uni ZAP XR Vector; 5'
adaptor sequence: 5' GAATTCGGCACGAG 3' 3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
BASE COUNT 126 a 108 c 109 g 91 t 1 others
ORIGIN

alignment scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2_COPY_178_186 x AA139849 ..

Align Seq 1/1 to: AA139849 from: 1 to: 449

1 LeuAlaGlyLysGluLeuCys 7
|||||||||||||||||||||
400 TTAGCAGGCAAGGAGCTTGC 420

Seq name: qb_est23:AI671776

Seq documentation block:
LOCUS AI671776 455 bp mRNA EST 16-DEC-1999
DEFINITION wa05h01.y1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2297233 3'
similar to contains Alu repetitive element;contains L1.t4 L1
repetitive element ;, mRNA sequence.
ACCESSION AI671776
VERSION AI671776.1 GI:4851507
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 455)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 628 Std Error: 0.00
Seq primer: 400P from Gibco.
FEATURES Location/Qualifiers
source 1..455
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2297233"
/clone_lib="NCI_CGAP_Kid11"
/lab_host="DH10B"
/note="Gram Kidney Vector pT7T3D (Pharmacia) with
a modified polylinker, Site 1: NotI, Site 2: EcoRI,
Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1422476 1423911, 1456007 1456775, and
1500552 1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo. "
BASE COUNT 93 a 88 c 94 g 181 t
ORIGIN

alignment scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2_COPY_178_186 x AI671776 ..

Align Seq 1/1 to: AI671776 from: 1 to: 455

1 LeuAlaGlyLysGluLeuCys 7
|||||||||||||||||||||
334 CTTGCAGGCAAGGAGCTTGGT 454

Seq name: qb_gss23:AZ256508

Seq documentation block:
LOCUS AZ256508 468 bp DNA GSS 26-JUL-2000
DEFINITION RPCI-23 122N21.TV RPCI-23 Mus musculus genomic clone RPCI-23 122N21
, DNA sequence.
ACCESSION AZ256508
VERSION AZ256508.1 GI:9460419
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 468)
AUTHORS Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinrot
,B., Levins,M., Mcqann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished (1999)
COMMENT Other GSSs: RPCI-23 122N21.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderinfoframe.htm)
or from Research Genetics (info@resgen.com). BAC end pairs:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 122 row: N column: 21
Seq primer: SP6
Class: BAC ends.
FEATURES Location/Qualifiers
source 1..468
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23:122N21"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Gram-Kidney/Brain, Vector: pBACe3.6, Site_1:
EcoRI, Site_2: EcoRI, Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoKI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies). "
BASE COUNT 150 a 106 c 87 g 125 t
ORIGIN

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0

```
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_178_186 x AZ256508/rev    .

 Align Seq 1/1  to reverse of: AZ256508  from: 1  to: 168

       1 LeuAlaGlyLysGluLeuCys 7
         |||||||||||||||||||
     208 CTAGCAGGGAAGGAACTATGC 188

seq_name: gb_est78:BE757615

seq_documentation_block:
LOCUS        BE757615       539 bp     mRNA             EST       15 SEP 2000
DEFINITION   21210% MARC 2B0V Bos taurus cDNA 5', mRNA sequence.
ACCESSION    BE757615
VERSION      BE757615.1  GI:10171607
KEYWORDS     EST.
SOURCE       cow.
  ORGANISM   Bos taurus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
             Bovidae; Bovinae; Bos.
REFERENCE    1  (bases 1 to 539)
  AUTHORS    Smith,T.P.L., Casas,E., Stone,R.T., Beaton,M.P., Grosse,W.M.,
             Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Kohrer,G.A., Laegreid
             ,W.W. and Keele,J.W.
  TITLE      Design and use of four pooled tissue normalized cDNA libraries for
             EST discovery in cattle
  JOURNAL    Unpublished (2000)
COMMENT      Contact: Smith TPL
             USDA, ARS, US Meat Animal Research Center
             PO Box 166  Clay Center  NE 68933-0166  USA
             Tel: 402 762 4366
             Fax: 402 762 4390
             Email: smith@email.marc.usda.gov
             Single pass sequencing. Bases called and alt_trimmed with phred
             v0.980904.e. Vector identified by cross_match with the -minscore 18
             and -minmatch 12 options.
             PCR Primers
             FORWARD: AGGAAACAGCTATGACCAT
             BACKWARD: GTTTTCCCAGTCACGACG
             Plate: 64  row: O  column: 13
             Seq primer: ATTIAGGIGACACTATAG.
FEATURES             Location/Qualifiers
     source          1..539
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 2B0V"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vectors: pCMV SPORT6, Site_1: XbaI, Site_2: XhoI,
                     Library made from pooled tissue from testis, thymus,
                     semitendonosus muscle, longissimus muscle, pancreas,
                     adrenal, and endometrium."
BASE COUNT       98 a    168 c    193 g     80 t
ORIGIN


alignment_scores:
             Quality:     7.00              Length:       7
               Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_178_186 x BE757615    ..

 Align Seq 1/1  to: BE757615  from: 1  to: 539

       3 GlyLysGluLeuCysAlaMet 9
         |||||||||||||||||||||
     290 GGCAAGGAGCTGTGTGCCATG 310

seq_name: gb_gss28:AZ560668

seq_documentation_block:
LOCUS        AZ560668       584 bp    DNA              GSS       20-NOV-2000
DEFINITION   RPCI-23-216E12.TV RPCI-23 Mus musculus genomic clone RPCI-23-216E12
             , DNA sequence.
ACCESSION    AZ560668
VERSION      AZ560668.1  GI:11240488
KEYWORDS     GSS.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 584)
  AUTHORS    Zhao,S., Nierman,W., Feldblyum,I., Malek,J., Shatsman,S., Akinre
             ,b., Levins,M., Mcgann,S., Isoqaye,G., Geer,K., Kriel,M., de Jong,P.
             and Fraser,C.M.
  TITLE      Mouse BAC End Sequences from Library RPCI-23
  JOURNAL    Unpublished (1999)
COMMENT      Other_GSSs: RPCI-23-216E12.TJ
             Contact: Shaying Zhao
             Department of Eukaryotic Genomics
             The Institute for Genomic Research
             9712 Medical Center Dr., Rockville, MD 20850, USA
             Tel: 301 838 0200
             Fax: 301 838 0208
             Email: szhao@tigr.org
             Clones are derived from the mouse BAC library RPCI-23. For BAC
             library availability, please contact Pieter de Jong
             (pieter@dejong.med.buffalo.edu). Clones may be purchased from
             BACPAC R...  ...s (http://bacpac.med.buffalo.edu/orderinfoframe.htm)
             or from Research Genetics (info@resgen.com). BAC end page:
             http://...ssg.tigr.org/db/bac_ends/mouse/bac_end_intro.html
             Plate: 216  row: E  column: 12
             Seq primer: T7
             Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..584
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-216E12"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Vector: Bluescript; Site_1: ...; Site_2: ...;
                     F...ST; Site_2: ...ST; Female C57BL/6J ... Kidney, adult
                     brain genomic DNA was isolated and partially digested
                     with a combination of EcoRI and EcoRI Methylase.  Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI sites.  The ligation products were transformed into
                     DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT      186 a    141 c    113 g    154 t
ORIGIN


alignment_scores:
             Quality:     7.00              Length:       7
               Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_178_186 x AZ560668/rev    ..

 Align Seq 1/1  to reverse of: AZ560668  from: 1  to: 584

       1 LeuAlaGlyLysGluLeuCys 7
         |||||||||||||||||||
     185 CTAGCAGGGAAGGAACTATGC 165

seq_name: gb_est84:BE156352
```

```
seq documentation block:
LOCUS        BF156352       638 bp     mRNA            EST       23 FEB 2001
DEFINITION   tl51a04.y1 Suqano Kawakami zebrafish DRA Danio rerio cDNA clone
             3817231 5' similar to TR:O43264 O43264 CENTROMERE/KINETOCHORE
             PROTEIN ZW10 HOMOLOG. ;, mRNA sequence.
ACCESSION    BF156352
VERSION      BF156352.1  GI:11051535
KEYWORDS     EST.
SOURCE       zebrafish.
  ORGANISM   Danio rerio
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
             Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE    1  (bases 1 to 638)
  AUTHORS    Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
             ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
             ,K., Stoptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
             Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
             Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
             and Wilson,R.
  TITLE      WashU Zebrafish EST Project 1998
  JOURNAL    Unpublished (1998)
COMMENT      Contact: Stephen L. Johnson
             Washington University School of Medicine
             4444 Forest Park Parkway, Box 8501, St. Louis, Mo 63108, USA
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: zbrafish@watson.wustl.edu
             Library constructed by Dr. Sumio Sugano and Dr. Keiichi Kawakami DNA
             Sequencing by: Washington University Genome Sequencing Center Clone
             distribution information can be found through the I.M.A.G.E.
             Consortium/LLNL, send email to: info@image.llnl.gov
             Seq primer: T3 ET from Amersham
             High quality sequence stop: 526.
FEATURES             Location/Qualifiers
     source          1..638
                     /organism="Danio rerio"
                     /strain="Ab"
                     /db_xref="taxon:7955"
                     /clone="3817231"
                     /clone_lib="Sugano Kawakami zebrafish DRA"
                     /sex="mixed (one male and one female, including
                     unfertilized eggs)"
                     /dev_stage="adult"
                     /lab_host="DH10B (phage resistant)"
                     /note="Vector: pME18S-FL3; Site_1: BcaBI (CACTGTGTG);
                     Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
                     with an oligo(dT) primer [AIGTGGCCTTTTTTTTTTTTTTTT];
                     double stranded cDNA was ligated to a  DraIII adaptor
                     [TGTTGGCCTACTGG], digested and cloned into disrupted BcaBI
                     sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
                     CACCATGTG). XhoI should be used to isolate the cDNA
                     insert.  Size selection was performed to exclude fragments
                     <1.5kb.  Library constructed by Dr. Sumio Sugano
                     (University of Tokyo Institute of Medical Science) and
                     kindly donated by Dr. Keiichi Kawakami.  Custom primers for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA.  "
BASE COUNT       218 a    128 c    153 g    139 t
ORIGIN


alignment scores:
            Quality:     7.00            Length:       7
              Ratio:    1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY-178-186 x BF156352  ..

 Align seq 1/1  to: BF156352  from: 1  to: 638

       1 LeuAlaGlyLysGluLeuCys 7
 .

         .
```

```
                  1 ||| | |||||||||||||
      434 CTTGCACGAAAACAATTAIGT 454

seq name: gb:gss27:AZ470308

seq documentation block:
LOCUS        AZ470308       653 bp    DNA              GSS       04-OCT-2000
DEFINITION   IM0284O13F Mouse 10kb plasmid UHGCIM library Mus musculus genomic
             clone UHGCIM0284O13 F, DNA sequence.
ACCESSION    AZ470308
VERSION      AZ470308.1  GI:10628433
KEYWORDS     GSS.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 653)
  AUTHORS    Dunn,D., Aoyagi,A., Barber,M., Boacorn,I., Duval,B., Bamil,C.,
             Islam,H., Longacre,S., Mahmood,M., Meenen,E., Pedersen,T., Reilly
             ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
             and Wright,D.,Weiss,R.
  TITLE      Mouse whole genome scaffolding with paired end reads from 10kb
             plasmid inserts
  JOURNAL    Unpublished (2000)
COMMENT      Contact: Robert R. Weiss
             University of Utah Genome Center
             University of Utah
             Rm. 408, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
             84112, USA
             Tel: 801 585 5606
             Fax: 801 585 7177
             Email: ddunn@genetics.utah.edu
             Insert Length: 10000  Std Error: 0.00
             Plate: 0284  row: O  column: 13
             Seq primer: CGTIGIAAAACGACGGCCAGT
             Class: plasmid ends
             High quality sequence stop: 653.
FEATURES             Location/Qualifiers
     source          1..653
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UHGCIM0284O13"
                     /clone_lib="Mouse 10kb plasmid UHGCIM library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10 Gold, tl resistant, F'"
                     /note="Vector: PWO42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.008 inch orifice at constant velocity. The sheared DNA
                     was blunt end repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5  to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWO42 (gi|4742114|gb|AF129072.1), a copy number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       155 a    153 c    176 g    169 t
ORIGIN


alignment scores:
            Quality:     7.00            Length:       7
              Ratio:    1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000
```

```
alignment_block:
 US-09-126-945B-2_COPY_178_186 x AZ470308/rev   ..

 Align seq 1/1  to: reverse of: AZ470308  from: 1  to: 654

       1 LeuAlaGlyLysGluLeuCys 7
         ||||||||||||||||||||
     100 CTAGCTGGGAAGGAACTCTGT 80

Seq_name: gb_est85:BF243620

Seq_documentation_block:
LOCUS       BF243620      853 bp    mRNA            EST       14-NOV-2000
DEFINITION  601879169F1 NIH_MGC_55 Homo sapiens cDNA clone IMAGE:4107760 5',
            mRNA sequence.
ACCESSION   BF243620
VERSION     BF243620.1  GI:11157549
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 853)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM989  row: j  column: 17
            High quality sequence stop: 622
FEATURES             Location/Qualifiers
     source          1..853
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4107760"
                     /clone_lib="NIH_MGC_55"
                     /tissue_type="from acute myelogenous leukemia"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: bone marrow; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgacctcgagcc);
                     Site_2: SfiI (ggccattatggccc
                     ); Double-stranded cDNA was prepared from cell line RNA.
                     5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C, G, or T).  Average
                     insert size 1.65 Kb (range 0.5-4.0 Kb).  14/15 colonies
                     contained inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      198 a    232 c    248 g    184 t      1 others
ORIGIN


alignment_scores:
            Quality:     7.00             Length:       7
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_178_186 x BF243620   ..

 Align seq 1/1  to: BF243620  from: 1  to: 853

       2 AlaGlyLysGluLeuCysAla 8
         ||||||||||||||||||||

     808 GCGGGCAAGGAACTTTGTGCC 828

Seq_name: gb_est85:BF579116

Seq_documentation_block:
LOCUS       BF579116      2064 bp    mRNA            EST       12-DEC-2000
DEFINITION  602860563F1 NCI_CGAP_CO24 Mus musculus cDNA clone IMAGE:4215961 5',
            mRNA sequence.
ACCESSION   BF579116
VERSION     BF579116.1  GI:11652828
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2064)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9792  row: d  column: 14
            High quality sequence stop: 645.
FEATURES             Location/Qualifiers
     source          1..2064
                     /organism="Mus musculus"
                     /strain="BVP/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4215961"
                     /clone_lib="NCI_CGAP_CO24"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: colon; Vector: pCMV SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.6 Kb.  Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      624 a    621 c    620 g    199 t
ORIGIN


alignment_scores:
            Quality:     7.00             Length:       7
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_178_186 x BF579116   ..

 Align seq 1/1  to: BF579116  from: 1  to: 2064

       8 GlyLysGluLeuCysAlaMet 9
         ||||||||||||||||||||
     493 GGTAAGGAGCTGTGCGCCATG 513

Seq_name: gb_est18:AI818448

Seq_documentation_block:
LOCUS       AI818448      106 bp    mRNA            EST       18-MAR-1999
DEFINITION  ta78e07.xal NCI_CGAP_Bm2 Homo sapiens cDNA clone IMAGE:2138287 3',
            similar to gb:Z26876 60S RIBOSOMAL PROTEIN L38 (HUMAN);contains
            element OFR repetitive element ;, mRNA sequence.
ACCESSION   AI818448
VERSION     AI818448.1  GI:4034228
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 106)
  AUTHORS   NCI CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Herbert Morse, M.D., Michael R. Emmert-Buck,
            M.D., Ph.D.
             cDNA Library Preparation: David S. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www.bio.llnl.gov/ncicgap/ncicgap.html

            Trace considered overall poor quality
            Insert Length: 338  Std Error: 0.00
            Seq primer: 400P from Gibco
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1..106
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2050200"
                     /clone_lib="NCI_CGAP_HSC2"
                     /tissue_type="stem cell, 34+/38+"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: bone marrow; Vector: pAMP1; mRNA made from
                     bone marrow, stem cells 34+/38+, cDNA made by oligo dT
                     priming.  Directionally cloned.  Size selected on agarose
                     gel, average insert size 400 bp.  Primary library,
                     non-amplified.  "
BASE COUNT        38 a      19 c      24 g      25 t
ORIGIN


Alignment scores:
            Quality:     6.00              Length:       6
              Ratio:    1.000               Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

Alignment block:
 US-09-126-945B-2 copy 178_186 x AI418348  ..

 Align seq 1/1  to: AI418348  from: 1  to: 106

        1 LeuAlaGlyLysGluLeu 6
          ||||||||||||||||||
       44 TTGGCAGGGAAAGAACTG 60
```

```
CM of: US-09-126-945B-2_COPY_192_200 to: GenEmbl:*  -out_format pts

Date: Sep 14, 2001  9:38 PM

About: Results were produced by the GenCore software, version 4.5.
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n model -DEV=xlp
 -Q "/cgn2_1/ftptmp/spouse/OSP01_6645/1_out_11[illegible]/1[illegible]/app_query_fast_1_1.qry"
 -DB GenEmbl -QFMT fastap -SUFFIX olip2n.rge  GAPOP 4.500
 -GAPEXT 0.050 -MINMATCH 0.100 -LOOPCL 0.000 -LOOPEXT 0.000
 -QGAPOP 4.500 -QGAPEXT 0.050 -XGAPOP=60.000 -XGAPEXT 60.000
 -FGAPOP 6.000 -FGAPEXT 7.000  YGAPOP=60.000 -YGAPEXT 60.000
 -DELOP 6.000  DELEXT 7.000 -START 1 -MATRIX oligo
 -TRANS human40.cdi -LIST=45 -BOCALIGN=200 -THR_SCORE quality
 -THR_MIN 1 -ALIGN=15 -MODE=LOCAL -OUTFMT pts -NORM=ext  MINLEN 0
 -MAXLEN 2000000000 -USER=US09126945_3CGN1_1_7594  NCPU 6
 -ICPU 3  LONGLOG -NO_XLPXY -WAIT -THREADS 1

Search information block:
Query: US-09-126-945B-2_COPY_192_200
Query length: 9
Database: GenEmbl:*
Database sequences: 1344157
Database length: -866060004
Search time (sec): 6503.700000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT 60.000


score_list:
Sequence                Strd Orig    Score    Pct Sc Ln  %Id  Overlap
qb_pat1:AX048148          +    9.00   149.14   4.53       1894  ! AX048148 Sequence 1 from Paten
qb_pr4:AF071538           +    9.00   149.14   4.53       1894  ! AF071538 Homo sapiens Ets tran
qb_pr1:AB041549           +    9.00   149.12   4.54       1898  ! AB041549 Homo sapiens hSH mRNA
qb_pat1:AX048159          +    9.00   145.98   6.79       3317  ! AX048159 Sequence 12 from Pate
qb_pr6:AL157372           -    9.00   121.52   106.52     150664  ! AL157372 Homo sapiens mRNA; cDNA
qb_htg8:AC023218          -    8.00   111.25   583.93     76744  ! AC023218 Homo sapiens chromos
qb_htg24:CNS01DUS         -    8.00   106.41   1.1e+03    181647  ! AL133367 Homo sapiens chromo
qb_htg24:CNS01DXG         -    8.00   106.11   1.1e+03    191562  ! AL139300 Homo sapiens chromo
qb_pr6:AL139123           -    8.00   105.60   1.2e+03    209527  ! AL139123 Human DNA sequence
qb_ba2:AP003011           -    8.00   102.77   1.7e+03    346510  ! AP003011 Mesorhizobium loti
qb_pat2:AX069957          +    7.00   123.99   114.02     384  ! AX069957 Sequence 429 from Pate
qb_bal:AF013517           -    7.00   120.68   174.38     692  ! AF013517 Uncultivated soil bact
qb_pat2:AX081564          -    7.00   118.60   227.59     1001  ! AX081564 Sequence 69 from Pate
qb_pl1:AF047034           -    7.00   117.84   250.92     1146  ! AF047034 Zea mays putative per
qb_pr9:HSM800884          +    7.00   113.86   418.00     2325  ! AL110227 Homo sapiens mRNA; cD
qb_pr10:HSU33953          -    7.00   113.86   418.00     2325  ! U33953 Human Down Syndrome re
qb_ro1:MMPEA3             -    7.00   113.66   428.97     2419  ! X63190 M.musculus PEA3 mRNA,
qb_ba3:HIUI3961           +    7.00   113.04   464.12     2688  ! UI3961 Haemophilus influenzae
qb_ba3:HIU60833           +    7.00   112.86   475.82     2726  ! U60833 Haemophilus influenzae
qb_ro1:MMWRSA             +    7.00   112.84   476.27     2786  ! X69656 M.musculus (clone S5) W
qb_ba3:HIU60834           +    7.00   112.77   480.82     2824  ! U60834 Haemophilus influenzae
qb_pat1:A38690            +    7.00   112.52   496.21     2949  ! A38690 Sequence 1 from Patent
qb_pat1:AR102148          +    7.00   112.52   496.21     2949  ! AR102148 Sequence 1 from patei
qb_pat1:A38694            +    7.00   112.52   496.33     2950  ! A38694 Sequence 5 from Patent
qb_pat1:AR102150          +    7.00   112.52   496.33     2950  ! AR102150 Sequence 5 from patei
qb_pat1:A38696            +    7.00   112.48   499.24     2974  ! A38696 Sequence 7 from Patent
qb_pat1:AR102151          +    7.00   112.48   499.24     2974  ! AR102151 Sequence 7 from patei
qb_pat1:A38692            +    7.00   112.46   500.45     2984  ! A38692 Sequence 3 from Patent
qb_pat1:AR102149          +    7.00   112.46   500.45     2984  ! AR102149 Sequence 3 from patei
qb_pat1:A38744            +    7.00   112.45   500.81     2987  ! A38744 Sequence 55 from Patent
qb_pat1:AR102157          +    7.00   112.45   500.81     2987  ! AR102157 Sequence 55 from patei
qb_pat1:A38698            +    7.00   112.45   501.05     2989  ! A38698 Sequence 9 from Patent
qb_pat1:AR102152          +    7.00   112.45   501.05     2989  ! AR102152 Sequence 9 from patei
qb_pr1:AB051468           +    7.00   109.40   740.33     5135  ! AB051468 Homo sapiens mRNA for
qb_ov:AF284045            -    7.00   109.35   745.94     5189  ! AF284045 Melanostrammus aesleti
qb_ba3:HIU60832           +    7.00   109.29   750.81     5236  ! U60832 Haemophilus influenzae
qb_ba3:HIU60831           +    7.00   109.08   771.59     5438  ! U60831 Haemophilus influenzae
qb_ro2:EN3334120          +    7.00   108.72   808.41     5803  ! AJ334120 Rattus norvegicus mRN
qb_bal:AF067776           +    7.00   108.15   866.07     6413  ! AF067776 Abiotrophia defectiv

qb_pr1:AB037715           +    7.00   102.81   908.14     6816  ! AB037715 H.sa. Sapiens mRNA
qb_ro1:AF133401           +    7.00   102.36   962.78     7391  ! AF133401 Rattus norvegicus
qb_ba3:U32774             +    7.00   105.57   1.2e+03    10151  ! U32774 Haemophilus influen
qb_htg4:AC014010          +    7.00   102.94   1.7e+03    16216  ! AC014010 Treponema pallidum
qb_in3:CE2C168            +    7.00   100.21   2.4e+03    26307  ! Z70832 Caenorhabditis elega
qb_ba3:MTCY27             +    7.00    99.96   2.5e+03    27548  ! Z95208 Mycobacterium tuber

seq_name: qb_pat1:AX048148

seq_documentation_block:
LOCUS       AX048148      1894 bp    DNA              PAT       15-DEC-2000
DEFINITION  Sequence 1 from Patent WO0070092.
ACCESSION   AX048148
VERSION     AX048148.1  GI:11876967
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo
REFERENCE   1  (bases 1 to 1894)
  AUTHORS   Kaufmann,J. and Xin,H.
  TITLE     Expression of ets domain proteins in cancer
  JOURNAL   Patent: WO 0070092-A 1 23-NOV-2000;
            Chiron Corporation (US)
FEATURES             Location/Qualifiers
     source          1..1894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             416..1423
                     /note="Human JKE1S."
                     /codon_start=1
                     /protein_id="CAC19002.1"
                     /db_xref="GI:11876968"
                     /translation="MSAPPSLLLLLLVLPLALLLLLVLPLPLLAAA[illegible]
                     PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAPGASSREEPPEEFEQPPVTLSLAPA
                     [illegible]
                     WPPLVSFPMPFALPLLAPLLCAMSLPLPWPTPLVLHARLLIKESAAWMPER
                     TSPGAIHYPASISPESWTDSEVDSSPSPHLWPLPLLLRPRSYCPPPWSNEER
                     [illegible]
                     1"
BASE COUNT       468 a     653 c     571 g     302 t
ORIGIN


alignment_scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_192_200 x AX048148  ..

 Align seq 1/1  to: AX048148  from: 1  to: 1894

        1 ArgGlnArgSerProLeuGlyGlyAsp 9
          ||||||||||||||||||||||||||
      989 CGCCAGCGCCTCGGCCCTGGGTGGGGAT 1015

seq_name: qb_pr4:AF071538

seq_documentation_block:
LOCUS       AF071538      1894 bp    mRNA             PRI       16-JAN-2000
DEFINITION  Homo sapiens Ets transcription factor PDEF (PDEF) mRNA, complete
            cds.
ACCESSION   AF071538
VERSION     AF071538.1  GI:4007417
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo
REFERENCE   1  (bases 1 to 1894)
  AUTHORS   [illegible]
```

```
            Boltax,I., Crall,F., Dube,A., Weiss,A., Brown,L., Quinn,G., Kao,K.,
            Endress,G., Kunsch,C. and Libermann,T.A.
  TITLE     PDEF, a novel prostate epithelium specific ets transcription
            factor, interacts with the androgen receptor and activates
            prostate specific antigen gene expression
  JOURNAL   J. Biol. Chem. 275 (2), 1219-1225 (2000)
  MEDLINE   20092894
REFERENCE   2  (bases 1 to 1894)
  AUTHORS   Libermann,T.A., Oettgen,P., Kunsch,C. and Finger,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUN-1998) Medicine, Beth Israel Deaconess Medical
            Center, 330 Brookline Ave., Boston, MA 02215, USA
FEATURES             Location/Qualifiers
     source          1..1894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6p21.3"
                     /tissue_type="prostate"
     gene            1..1894
                     /gene="PDEF"
     CDS             416..1423
                     /gene="PDEF"
                     /note="prostate epithelium specific"
                     /codon_start=1
                     /product="Ets transcription factor PDEF"
                     /protein_id="AAC52606.1"
                     /db_xref="GI:4007418"
                     /translation="MGSASPGLSSVSPSHELLPPDIVSPTGLEKAAAGAVGLERRDWS
                     PSPPATPEQGLSAFYLSYFGMLYPEQSSWAAKAGSASSFEEPPEEFEQCPVTDSQAPA
                     GSLDLVPGGLELEERELEQVQSMVVGEVLKTETEACKELNLTADPMDWSGSSNVQKWLL
                     WTEHQYRLPPMGKAEQELAGKELGAMSEEQLRQKSFDGGFVLHAHLLDWKSAAWMKER
                     TSPGATHYCASTSEESWTDSEVDSSGSQGPTHLWQFSGELKELLERPESYGRFTRWLNKEK
                     GIFKTEDSAQVARLWGTRKNRPAMNYDKLSRSLRQYYKKGILEKLDLSGRLVYQFVHP
                     I"
BASE COUNT       468 a    654 c    571 g    302 t
ORIGIN


alignment scores:
            Quality:     9.00            Length:       9
              Ratio:    1.000              Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2_COPY_192_200 x AF071538    ..

 Align seq 1/1   to: AF071538   from: 1   to: 1894

        1 ArgGlnArgSerProLeuGlyGlyAsp 9
          ||||||||||||||||||||||||||||
      989 CGCCAGCGCTCGCCCCTGGGTGGGGAT 1015

seq name: qb_pri:AB031549

seq documentation block:
LOCUS       AB031549      1898 bp    mRNA            PRI       20-JAN-2000
DEFINITION  Homo sapiens PSE mRNA for prostate ets, complete cds.
ACCESSION   AB031549
VERSION     AB031549.1  GI:6721497
KEYWORDS    prostate ets.
SOURCE      Homo sapiens adult male prostate epithelium prostate cancer
            cell line:PC-3 cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Yamada,N., Tamai,Y., Miyamoto,H. and Nozaki,M.
  TITLE     Cloning and expression of the mouse Pse gene encoding a novel Ets
            family member
  JOURNAL   Gene 241 (2), 267-274 (2000)
  MEDLINE   20137504
REFERENCE   2  (bases 1 to 1898)
  AUTHORS   Nozaki,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-AUG-1999) to the DDBJ/EMBL/GenBank databases. Masami
            Nozaki, Osaka University, Institute for Microbial Diseases; 3-1,
            Yamadaoka, Suita, Osaka 565-0871, Japan
            (E-mail:mnozaki@biken.osaka-u.ac.jp, Tel:81-6-879-8438,
            Fax:81-6-879-8439)
FEATURES             Location/Qualifiers
     source          1..1898
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="PC-3"
                     /cell_type="prostate cancer"
                     /chromosome="6"
                     /dev_stage="adult"
                     /map="6p21.2-21.3"
                     /sex="male"
                     /tissue_type="prostate epithelium"
                     /note="vector:pUC18"
     gene            406..1413
                     /gene="PSE"
     CDS             406..1413
                     /gene="PSE"
                     /codon_start=1
                     /product="prostate ets"
                     /protein_id="BAA89543.1"
                     /db_xref="GI:6721498"
                     /translation="MGSASPGLSSVSPSHELLPPDIVSPTGLEKAAAGAVGLERRDWS
                     PSPPATPEQGLSAFYLSYFGMLYPEQSSWAAKAGSASSFEEPPEEFEQCPVTDSQAPA
                     GSLDLVPGGLELEERELEQVQSMVVGEVLKTETEACKELNLTADPMDWSGSSNVQKWLL
                     WTEHQYRLPPMGKAEQELAGKELGAMSEEQLRQKSFDGGFVLHAHLLDWKSAAWMKER
                     TSPGATHYCASTSEESWTDSEVDSSGSQGPTHLWQFSGELKELLERPESYGRFTRWLNKEK
                     GIFKTEDSAQVARLWGTRKNRPAMNYDKLSRSLRQYYKKGILEKLDLSGRLVYQFVHP
                     I"
BASE COUNT       380 a    651 c    567 g    300 t
ORIGIN


alignment scores:
            Quality:     9.00            Length:       9
              Ratio:    1.000              Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2_COPY_192_200 x AB031549    ..

 Align seq 1/1   to: AB031549   from: 1   to: 1898

        1 ArgGlnArgSerProLeuGlyGlyAsp 9
          ||||||||||||||||||||||||||||
      979 CGCCAGCGCTCGCCCCTGGGTGGGGAT 1005

seq name: qb_pat1:AX048159

seq documentation block:
LOCUS       AX048159      3317 bp    DNA             PAT       15-DEC-2000
DEFINITION  Sequence 12 from Patent WO0070092.
ACCESSION   AX048159
VERSION     AX048159.1  GI:11876975
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3317)
  AUTHORS   Kaufmann,J. and Xin,H.
  TITLE     Expression of ets-domain proteins in cancer
  JOURNAL   Patent: WO 0070092-A 12-24-NOV-2000;
            Chiron Corporation (US)
FEATURES             Location/Qualifiers
     source          1..3317
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

```
                     /note="GSEF encoding sequence with promoter (Pias  2A-2B)"
     promoter        1. .1380
                     /note="Promoter"
     CDS             1815. .2822
                     /note="GSEF coding sequence"
                     /codon_start 1
                     /protein_id "CAC10004.1"
                     /db_xref "GI:11876976"
                     /translation "MGSASPGLSSVSPSHLLLLPFFLVSPFGLEKAAAGAVGLERRMWS
                     PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKANGASSREEDPEEPEQGPVFTSQAPA
                     GSLDLVPGGLPLEEHSLEQVQSMVVGPVLKGITACKLLNLTACFMLWSPGNVGKWLL
                     WTEHQYELPPMGKAPQPLACKPLGAMSPQPEWQMSPDGCPVTHAHCPLWPSAAWMVGL
                     ISPGATHYCASISPFSWTLDSPVDSSGSQGPDHPWQPLKELLPPHSYGPFPWLNSPP
                     GIFKIEDSAQVARTWGTPPNPPAMNYQKTGPQRVYPPGGTGSPPDTSGPVVYGPVHP
                     I"
BASE COUNT       710 a   1026 c    979 g    611 t
ORIGIN


alignment_scores:
             Quality:    9.00              Length:       9
               Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_192_200 x AX048159  ..

 Align seq 1/1  to: AX048159  from: 1  to: 3317

       1 ArgGlnArgSerProLeuGlyGlyAsp 9
         ||||||||||||||||||||||||||||
    2388 CGCCAGCGCTCGCCCCTGGGTGGAGAT 2414

Seq_name: gb_pr6:AL157372

Seq_documentation_block:
LOCUS        AL157372   150660 bp    DNA             PRI       29-NOV-2000
DEFINITION   Human DNA sequence from clone RP11-475E1 on chromosome 6, complete
             sequence.
ACCESSION    AL157372
VERSION      AL157372.18  GI:11493236
KEYWORDS     HTG.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 150660)
  AUTHORS    Skuce,C.
  TITLE      Direct Submission
  JOURNAL    Submitted (29-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
             CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
             requests: clonerequest@sanger.ac.uk
COMMENT      On Nov 30, 2000 this sequence version replaced gi:11342772.
             During sequence assembly data is compared from overlapping clones.
             Where differences are found these are annotated as variations
             together with a note of the overlapping clone name. Note that the
             variation annotation may not be found in the sequence submission
             corresponding to the overlapping clone, as we submit sequences with
             only a small overlap as described above.
             This sequence has been finished according to sequence map criteria
             as follows.  An attempt is made to resolve all sequencing problems,
             such as compressions and repeats, but not necessarily within known
             annotated repeat sequence elements.  Where the sequence is
             ambiguous, there is an annotation using the 'unsure' feature key.
             The following abbreviations are used to associate primary accession
             numbers given in the feature table with their source databases:
             Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
             on the WORMPEP database can be found at
             http://www.sanger.ac.uk/Projects/C_elegans/wormpep HTG sequence
             was generated from part of bacterial clone contigs of human
             chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
             Group.  Further information can be found at
             http://www.sanger.ac.uk/HGP/Chr6/
             RP11-475E1 is from the library RPCI 11.2 constructed at the Roswell
             Park Cancer Institute by the group of Pieter de Jong. For further
             details see http://bacpac.med.buffalo.edu/
             VECTOR: pBACe3.6
             IMPORTANT: This sequence is not the entire insert of clone
             RP11-475E1. It may be shorter because we sequence overlapping
             sections only once, except for a 100 base overlap.
             The true left end of clone RP11-475E1 is at 1 in this sequence. The
             true left end of clone RP3-391D22 is at 150561 in this sequence.
             The true right end of clone RP11-802L10 is at 113044 in this
             sequence.
FEATURES             Location/Qualifiers
     source          1. .150660
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-475E1"
                     /clone_lib="RPCI 11.2"
BASE COUNT    35470 a   38867 c   38740 g   37583 t
ORIGIN


alignment_scores:
             Quality:    9.00              Length:       9
               Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_192_200 x AL157372/rev  ..

 Align seq 1/1  to reverse of: AL157372  from: 1  to: 150660

       1 ArgGlnArgSerProLeuGlyGlyAsp 9
         ||||||||||||||||||||||||||||
  126329 CGCCAGCGCTCGCCCCTGGGTGGGGAT 126303

Seq_name: gb_htg8:AC024218

Seq_documentation_block:
LOCUS        AC024218    76734 bp    DNA             HTG       14-JUL-2000
DEFINITION   Homo sapiens chromosome 14 clone RP11-105M4 map 14, LOW PASS
             SEQUENCE SAMPLING.
ACCESSION    AC024218
VERSION      AC024218.2  GI:9165238
KEYWORDS     HTG; HTGS_PHASE0.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 76734)
  AUTHORS    Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE      Homo sapiens chromosome 14, clone RP11-105M4
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 76734)
  AUTHORS    Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,B., Allen,N.,
             Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
             Bolotarsky,I., Bookhalter,B., Brown,A., Burkett,G., Castle,A.,
             Chaopei,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,F.,
             DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenester,J.,
             Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Gallagan,J.,
             Gardyna,S., Grant,G., Hagos,B., Hanford,A., Hotton,I.,
             Howland,J.C., Johnson,R., Jones,C., Kamal,L., Karatas,A., Klein,J.,
             Landers,I., Lebowsky,J., Levine,R., Liew,C., Liu,C., Locke,E.,
             Macdonald,P., Margolis,N., McEwan,P., McGurk,A., McKernan,K.,
             McPherson,R., Meldrim,J., Meneses,L., Morrow,J., Naylor,J.,
             Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,I.M., Peterson,K.,
             Pierre,N., Pisani,C., Pollard,V., Raymond,C., Riley,K., Rittmann,I.,
             Roy,A., Santos,R., Severy,P., Spencer,B., Stanton-Thomann,N.,
             Steffen,N., Subramanian,A., Tilanus,J., Tsonis,..., Theodore,J.,
             Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Yo,W.J.,
             Zimmer,A. and Zody,M.
  TITLE      Direct Submission
  JOURNAL    Submitted (19-FEB-2000) Whitehead Inst/MIT Center for Genome
```

```
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 14, 2000 this sequence version replaced gi:6957752.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
                        Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www.seq.wi.mit.edu
                Contact: sequence-submissions@genome.wi.mit.edu
                        Project Information
                Center project name: L6632
                Center clone name: 105_M_4
                        .
            * NOTE: This record contains 81 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low pass sequence sampling is useful for
            * identifying clones that may be gene rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      871: contig of 871 bp in length
            *      872  971: gap of      100 bp
            *      972     1848: contig of 867 bp in length
            *     1849  1948: gap of      100 bp
            *     1949     2804: contig of 866 bp in length
            *     2805  2904: gap of      100 bp
            *     2905     3746: contig of 842 bp in length
            *     3747  3846: gap of      100 bp
            *     3847     4667: contig of 821 bp in length
            *     4668  4767: gap of      100 bp
            *     4768     5627: contig of 860 bp in length
            *     5628  5727: gap of      100 bp
            *     5728     6551: contig of 824 bp in length
            *     6552  6651: gap of      100 bp
            *     6652     7527: contig of 876 bp in length
            *     7528  7627: gap of      100 bp
            *     7628     8459: contig of 832 bp in length
            *     8460  8559: gap of      100 bp
            *     8560     9444: contig of 885 bp in length
            *     9445  9544: gap of      100 bp
            *     9545    10368: contig of 824 bp in length
            *    10369 10468: gap of      100 bp
            *    10469    11325: contig of 857 bp in length
            *    11326 11425: gap of      100 bp
            *    11426    12274: contig of 849 bp in length
            *    12275 12374: gap of      100 bp
            *    12375    13223: contig of 849 bp in length
            *    13224 13323: gap of      100 bp
            *    13324    14163: contig of 840 bp in length
            *    14164 14263: gap of      100 bp
            *    14264    15057: contig of 794 bp in length
            *    15058 15157: gap of      100 bp
            *    15158    16018: contig of 861 bp in length
            *    16019 16118: gap of      100 bp
            *    16119    16964: contig of 846 bp in length
            *    16965 17064: gap of      100 bp
            *    17065    17904: contig of 840 bp in length
            *    17905 18004: gap of      100 bp
            *    18005    18858: contig of 854 bp in length
            *    18859 18958: gap of      100 bp
            *    18959    19781: contig of 823 bp in length
            *    19782 19881: gap of      100 bp
            *    19882    20729: contig of 848 bp in length
            *    20730 20829: gap of      100 bp
            *    20830    21737: contig of 908 bp in length
            *    21738 21837: gap of      100 bp
            *    21838    22684: contig of 847 bp in length
            *    22685 22784: gap of      100 bp
            *    22785    23640: contig of 856 bp in length
            *    23641 23740: gap of      100 bp
            *    23741    24566: contig of 826 bp in length
            *    24567 24666: gap of      100 bp
            *    24667    25519: contig of 853 bp in length
            *    25520 25619: gap of      100 bp
            *    25620    26478: contig of 859 bp in length
            *    26479 26578: gap of      100 bp
            *    26579    27434: contig of 856 bp in length
            *    27435 27534: gap of      100 bp
            *    27535    28406: contig of 872 bp in length
            *    28407 28506: gap of      100 bp
            *    28507    29351: contig of 845 bp in length
            *    29352 29451: gap of      100 bp
            *    29452    30292: contig of 841 bp in length
            *    30293 30392: gap of      100 bp
            *    30393    31253: contig of 861 bp in length
            *    31254 31353: gap of      100 bp
            *    31354    32178: contig of 825 bp in length
            *    32179 32278: gap of      100 bp
            *    32279    33145: contig of 867 bp in length
            *    33146 33245: gap of      100 bp
            *    33246    34087: contig of 842 bp in length
            *    34088 34187: gap of      100 bp
            *    34188    35016: contig of 829 bp in length
            *    35017 35116: gap of      100 bp
            *    35117    35991: contig of 875 bp in length
            *    35992 36091: gap of      100 bp
            *    36092    36956: contig of 865 bp in length
            *    36957 37056: gap of      100 bp
            *    37057    37883: contig of 827 bp in length
            *    37884 37983: gap of      100 bp
            *    37984    38833: contig of 850 bp in length
            *    38834 38933: gap of      100 bp
            *    38934    39774: contig of 841 bp in length
            *    39775 39874: gap of      100 bp
            *    39875    40726: contig of 852 bp in length
            *    40727 40826: gap of      100 bp
            *    40827    41689: contig of 863 bp in length
            *    41690 41789: gap of      100 bp
            *    41790    42592: contig of 803 bp in length
            *    42593 42692: gap of      100 bp
            *    42693    43553: contig of 861 bp in length
            *    43554 43653: gap of      100 bp
            *    43654    44504: contig of 851 bp in length
            *    44505 44604: gap of      100 bp
            *    44605    45448: contig of 844 bp in length
            *    45449 45548: gap of      100 bp
            *    45549    46399: contig of 851 bp in length
            *    46400 46499: gap of      100 bp
            *    46500    47373: contig of 874 bp in length
            *    47374 47473: gap of      100 bp
            *    47474    48346: contig of 873 bp in length
            *    48347 48446: gap of      100 bp
            *    48447    49294: contig of 848 bp in length
            *    49295 49394: gap of      100 bp
            *    49395    50222: contig of 828 bp in length
            *    50223 50322: gap of      100 bp
            *    50323    51144: contig of 822 bp in length
            *    51145 51244: gap of      100 bp
            *    51245    52078: contig of 834 bp in length
            *    52079 52178: gap of      100 bp
            *    52179    53031: contig of 853 bp in length
            *    53032 53131: gap of      100 bp
            *    53132    53970: contig of 839 bp in length
            *    53971 54070: gap of      100 bp
            *    54071    54911: contig of 841 bp in length
            *    54912 55011: gap of      100 bp
            *    55012    55873: contig of 862 bp in length
            *    55874 55973: gap of      100 bp
            *    55974    56797: contig of 824 bp in length
            *    56798 56897: gap of      100 bp
            *    56898    57740: contig of 843 bp in length
```

```
            *    57741 57840: gap of       100 bp
            *    57841    58694: contig of 854 bp in length
            *    58695 58794: gap of       100 bp
            *    58795    59619: contig of 825 bp in length
            *    59620 59719: gap of       100 bp
            *    59720    60566: contig of 847 bp in length
            *    60567 60666: gap of       100 bp
            *    60667    61514: contig of 848 bp in length
            *    61515 61614: gap of       100 bp
            *    61615    62461: contig of 847 bp in length
            *    62462 62561: gap of       100 bp
            *    62562    63399: contig of 838 bp in length
            *    63400 63499: gap of       100 bp
            *    63500    64332: contig of 833 bp in length
            *    64333 64432: gap of       100 bp
            *    64433    65278: contig of 846 bp in length
            *    65279 65378: gap of       100 bp
            *    65379    66237: contig of 859 bp in length
            *    66238 66337: gap of       100 bp
            *    66338    67174: contig of 837 bp in length

alignment_scores:
            Quality:    8.00            Length:      8
              Ratio:   1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_192_200 x AC023218/rev  ..

 Align seq 1/1  to reverse of: AC023218  from: 1  to: 76744

       1 ArgGlnArgSerProLeuGlyGly 8
         ||||||||||||||||||||||||
   16741 AGGCAGAGAAGCCCGTTAGGAGGC 16718

seq_name: qb_htq24:CNS01DUS

seq_documentation_block:
LOCUS       CNS01DUS    181647 bp    DNA             HTG       15-DEC-2000
DEFINITION  Homo sapiens chromosome 14 clone R-600P24, *** SEQUENCING IN
            PROGRESS ***, in ordered pieces.
ACCESSION   AL133367
VERSION     AL133367.3  GI:11875945
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 181647)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-DEC-2000) Genoscope  Centre National de Sequencage :
            BP 191 91006 EVRY cedex  FRANCE (E-mail : Seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     On Dec 17, 2000 this sequence version replaced gi:7708222.
            -------------- Genome Center
            Center: Genoscope / Centre National de Sequencage
            Center code: GS
            Web site: http://www.genoscope.cns.fr/
            Contact: SeqRef@genoscope.cns.fr
            --------------
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continue. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage, etc. . . even if efforts are made to eliminate these
            contaminating sequences. The following BAC sequence is oriented
            from the T7 to the SP6 end.
            Upstream BAC (overlapping the T7 end) : R-894P9 (AC=AL109699)
            Downstream BAC (overlapping the SP6 end) : R-45P15 ----------
            Summary Statistics
            Assembly program: Phrap; version 2.0
            Quality coverage: 7.91x in Q20 bases; sum of contigs
            ------ - - -
            Overall quality chart :
            Range     :   bases
            0         :
            1  -  9   :      12
            10 - 19   :      85
            20 - 29   :     224
            30 - 39   :    1488
            40 - 49   :    6906
            50 - 59   :    8382
            60 - 69   :   14213
            70 - 79   :   26969
            80 - 89   :   50749
            90 - 99   :   72621
            ------  ---------- ---
            Percentage of bases with a quality value > 40 : 99 %.
            * NOTE: This is a 'working draft' sequence,
            * this sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES              location/qualifiers
     source           1..181647
                      /organism "Homo sapiens"
                      /db_xref "taxon:9606"
                      /chromosome "14"
                      /clone "R-600P24"
                      /clone_lib "RPCI-11"
     STS              269..467
                      /note "matching EMBL:AA133284
                      RHdb:RH66763
                      dbSTS:STS46692
                      Identified using the e-PCR software (G. Schuler)"
     STS              16397..16560
                      /note "matching EMBL:M16451
                      RHdb:RH59084
                      RHdb:RH17667
                      dbSTS:STS26301
                      Identified using the e-PCR software (G. Schuler)"
     STS              32622..32885
                      /note "matching EMBL:M80359
                      RHdb:RH44092
                      RHdb:RH1619
                      dbSTS:STS11392
                      Identified using the e-PCR software (G. Schuler)"
     STS              32780..32891
                      /note "matching EMBL:M80359
                      RHdb:RH53510
                      dbSTS:STSI8471
                      Identified using the e-PCR software (G. Schuler)"
     STS              109445..109613
                      /note "matching EMBL:R98192
                      RHdb:RH70198
                      dbSTS:STS50075
                      Identified using the e-PCR software (G. Schuler)"
BASE COUNT     51901 a   42095 c   41100 g   46551 t
ORIGIN


alignment_scores:
            Quality:    8.00            Length:      8
              Ratio:   1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_192_200 x CNS01DUS/rev  ..

 Align seq 1/1  to reverse of: CNS01DUS  from: 1  to: 181647

       1 ArgGlnArgSerProLeuGlyGly 8
         ||||||||||||||||||||||||
    5829 AGGCAGAGAAGCCCGTTAGGAGGC 5806
```

```
seq_name: gb htg24:CNS01DXG

seq_documentation block:
LOCUS        CNS01DXG    191562 bp    DNA              HTG       01 FEB 2001
DEFINITION   Homo sapiens chromosome 14 clone R-894P9, *** SEQUENCING IN
             PROGRESS ***, in ordered pieces
ACCESSION    AL149300
VERSION      AL149300.5  GI:12657318
KEYWORDS     HTG; HTGS PHASE2; HTGS DRAFT.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 191562)
  AUTHORS    Genoscope.
  TITLE      Direct Submission
  JOURNAL    Submitted (01 FEB 2001) Genoscope  Centre National de Sequencage :
             BP 191 91006 EVRY cedex   FRANCE (E-mail : seqref@genoscope.cns.fr
               Web : www.genoscope.cns.fr)
COMMENT      On Feb 2, 2001 this sequence version replaced gi:10046641.
                          Genome Center
             Center: Genoscope / Centre National de Sequencage
             Center code: GS
             Web site: http://www.genoscope.cns.fr/
             Contact: SeqRef@genoscope.cns.fr

             IMPORTANT: This sequence is unfinished and does not necessarily
             represent the correct sequence. Work on the sequence is in progress
             and the release of this data is based on the understanding that the
             sequence may change as work continue. The sequence may be
             contaminated with foreign sequence from E.coli, yeast, vector,
             phage, etc. , , even if efforts are made to eliminate these
             contaminating sequences. The following BAC sequence is oriented
             from the T7 to the SP6 end.
             Upstream BAC (overlapping the T7 end) : R-73M18 (AC:AL049840)
             Downstream BAC (overlapping the SP6 end) : R-600F24 ------------ .
             Summary Statistics
             Assembly program: Phrap; version 2.0
             Quality coverage: 5.59x in Q20 bases; sum of contigs

             Overall quality chart :
             Range     :    bases
             0         :      155
             1     9   :       46
             10    19  :      296
             20    29  :      785
             30    39  :     3226
             40    49  :    11952
             50    59  :    12895
             60    69  :    14908
             70    79  :    28779
             80    89  :    55088
             90    99  :    64442

             Percentage of bases with a quality value > 40 : 97 %.
             * NOTE: This is a 'working draft' sequence.
             * This sequence will be replaced
             * by the finished sequence as soon as it is available and
             * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..191562
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /clone="R-894P9"
                     /clone_lib="RPCI-11"
     STS             90815..91023
                     /note="matching EMBL:N39280
                     RHdb:RH69467
                     dbSTS:STS49345
                     Identified using the e-PCR software (G. Schuler)"
     STS             102099..102229
                     /note="matching EMBL:AA146734
                     RHdb:RH48201
                     dbSTS:STS41260
                     Identified using the e-PCR software (G. Schuler)"
     STS             121989..122215
                     /note="matching EMBL:N25712
                     RHdb:RH69464
                     dbSTS:STS49242
                     Identified using the e-PCR software (G. Schuler)"
     STS             145312..145500
                     /note="matching EMBL:AA134284
                     RHdb:RH66763
                     dbSTS:STS46692
                     Identified using the e-PCR software (G. Schuler)"
     STS             161437..161600
                     /note="matching EMBL:M16451
                     RHdb:RH69084
                     RHdb:RH17667
                     dbSTS:STS26301
                     Identified using the e-PCR software (G. Schuler)"
     STS             177679..177942
                     /note="matching EMBL:M80459
                     RHdb:RH44092
                     RHdb:RH1619
                     dbSTS:STS11492
                     Identified using the e-PCR software (G. Schuler)"
     STS             177847..177948
                     /note="matching EMBL:M80459
                     RHdb:RH53610
                     dbSTS:STS18471
                     Identified using the e-PCR software (G. Schuler)"
BASE COUNT     47966 a  49500 c  47840 g  46231 t      5 others
ORIGIN


alignment scores:
            Quality:     8.00              Length:       8
              Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 MS309 126 945P 2 CSLY 192 200 x CNS01DXG/rev   ..

 Align seq 1/1  to reverse of: CNS01DXG  from: 1  to: 191562

       1 AcgGinAraSerProGenGlyGly 8
         ||||.:.|:|||||||||||||||
  150871 AGGCAGAGAACCCGTIAGGAGGC 150848

seq_name: gb pri:AL139123

seq_documentation block:
LOCUS        AL139123   209527 bp    DNA              PRI       11 JAN 2001
DEFINITION   Human DNA sequence from clone RP11-500G10 on chromosome 10,
             complete sequence.
ACCESSION    AL139123
VERSION      AL139123.14  GI:10178428
KEYWORDS     H.G.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 209527)
  AUTHORS    Chapman,J.
  TITLE      Direct Submission
  JOURNAL    Submitted (11 JAN 2001) Sanger Centre, Hinxton, Cambridgeshire,
             CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
             requests: clonerequest@sanger.ac.uk
COMMENT      On Sep 15, 2020 this sequence version replaced gi:10129360.
             During sequence assembly data is compared from overlapping clones.
             Where differences are found these are annotated as variations
             together with a note of the overlapping clone name. Note that the
```

```
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP. Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep. This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 10, constructed by the Sanger Centre Chromosome 10
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10/
            This sequence is the entire insert of clone RP11-500G10 The true
            left end of clone RP11-435P9 is at 117409 in this sequence. The
            true right end of clone RP11-400F15 is at 107397 in this sequence.
            This sequence has been finished according to sequence map criteria
            as follows:  An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements.  Where the sequence is
            ambiguous, there is an annotation using the 'ensure' feature key.
            RP11-500G10 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong  For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1..209527
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-500G10"
                     /clone_lib="RPCI-11.2"
     repeat_region   1..223
                     /note="MLT1E repeat: matches 334, .568 of consensus"
     repeat_region   3819..3921
                     /note="MLT1D repeat: matches 118, .240 of consensus"
     repeat_region   3966..4010
                     /note="3 copies 15 mer 95% conserved"
     repeat_region   5339..5517
                     /note="MER4D repeat: matches 229, .407 of consensus"
     repeat_region   5516..5851
                     /note="MER4D repeat: matches 672, .981 of consensus"
     repeat_region   7061..7187
                     /note="MLT1H repeat: matches 413, .541 of consensus"
     repeat_region   7429..7761
                     /note="MER87 repeat: matches 194, .541 of consensus"
     repeat_region   7840..7890
                     /note="MLT1G repeat: matches 50, .101 of consensus"
     repeat_region   8081..8234
                     /note="MLT1B repeat: matches 1, .152 of consensus"
     repeat_region   8231..8320
                     /note="MLT1B repeat: matches 294, .384 of consensus"
     repeat_region   8721..8910
                     /note="MIR repeat: matches 19, .221 of consensus"
     repeat_region   9237..9502
                     /note="L1M1 repeat: matches 5644, .5846 of consensus"
     repeat_region   9635..9759
                     /note="L1MC3 repeat: matches 6128, .6246 of consensus"
     repeat_region   9760..10046
                     /note="AluSq repeat: matches 1, .291 of consensus"
     repeat_region   10047..11480
                     /note="L1MC3 repeat: matches 6246, .7740 of consensus"
     repeat_region   11772..12077
                     /note="AluSc repeat: matches 1, .309 of consensus"
     repeat_region   14021..14088
                     /note="34 copies 2 mer ac 76% conserved"
     repeat_region   15180..15248
                     /note="MER57A repeat: matches 99, .172 of consensus"
     repeat_region   15249..15554
                     /note="AluSq repeat: matches 1, .305 of consensus"
     repeat_region   15555..15923
                     /note="MER57A repeat: matches 1, .363 of consensus"
     repeat_region   16008..16113
                     /note="MIR repeat: matches 77, .185 of consensus"
     repeat_region   16236..16319
                     /note="MIR repeat: matches 62, .147 of consensus"
     repeat_region   17029..17096
                     /note="L1MB3 repeat: matches 5544, .5609 of consensus"
     repeat_region   17097..17397
                     /note="AluJb repeat: matches 6, .311 of consensus"
     repeat_region   17398..17958
                     /note="L1MB3 repeat: matches 5639, .6198 of consensus"
     repeat_region   18372..18680
                     /note="AluSq repeat: matches 1, .309 of consensus"
     repeat_region   19521..19577
                     /note="3 copies 19 mer 84% conserved"
     repeat_region   20031..20187
                     /note="MER5A repeat: matches 1, .157 of consensus"
     repeat_region   20258..20492
                     /note="3 copies 45 mer 80% conserved"
     repeat_region   20442..20522
                     /note="MER5A repeat: matches 108, .185 of consensus"
     repeat_region   21612..21990
                     /note="MER1B repeat: matches 1, .397 of consensus"
     repeat_region   23673..23805
                     /note="MER4C repeat: matches 281, .438 of consensus"
     repeat_region   23933..24314
                     /note="MER4A repeat: matches 1, .449 of consensus"
     repeat_region   24456..24894
                     /note="MLT1C repeat: matches 1, .466 of consensus"
     repeat_region   27442..27661
                     /note="5 copies 44 mer 74% conserved"
     repeat_region   27444..27663
                     /note="110 copies 2 mer tg 72% conserved"
     repeat_region   27658..27961
                     /note="16 copies 19 mer 56% conserved"
     repeat_region   27664..27803
                     /note="7 copies 20 mer 69% conserved"
     repeat_region   27670..27984
                     /note="7 copies 45 mer 67% conserved"
     repeat_region   27671..27751
                     /note="9 copies 9 mer gtgtgtgat 67% conserved"
     repeat_region   27678..27809
                     /note="22 copies 6 mer gtgtgt 68% conserved"
     repeat_region   27805..27956
                     /note="4 copies 38 mer 69% conserved"
     repeat_region   27808..27933
                     /note="7 copies 18 mer 69% conserved"
     repeat_region   27811..27930
                     /note="6 copies 20 mer 76% conserved"
     repeat_region   27859..27960
                     /note="6 copies 17 mer 71% conserved"
     repeat_region   27944..28151
                     /note="52 copies 4 mer gtgt 77% conserved"
     repeat_region   27949..28148
                     /note="10 copies 20 mer 79% conserved"
     repeat_region   27957..28146
                     /note="5 copies 38 mer 79% conserved"
     repeat_region   29059..29167
                     /note="MIR repeat: matches 78, .196 of consensus"
     repeat_region   31367..31564
                     /note="99 copies 2 mer gg 63% conserved"
     repeat_region   31372..31569
                     /note="11 copies 18 mer 62% conserved"
     repeat_region   31477..31571
                     /note="5 copies 19 mer 91% conserved"
     repeat_region   32001..32140
                     /note="7 copies 20 mer 72% conserved"
     repeat_region   32027..32292
                     /note="7 copies 38 mer 68% conserved"
     repeat_region   32096..32185
                     /note="5 copies 18 mer 74% conserved"
     repeat_region   32099..32215
                     /note="13 copies 9 mer tgtgtgcgc 68% conserved"
     repeat_region   32102..32191
                     /note="2 copies 45 mer 96% conserved"
     repeat_region   32152..32331
```

```
                     /note "9 copies 20 mer 75% conserved"
     repeat_region   32261. .32341
                     /note "9 copies 9 mer 100% conserved"
     repeat_region   32421. .32719
                     /note "AluSx repeat: matches 1. .309 of consensus"
     misc_feature    32714. .32739
                     /note "Sequence 1 is not 1 to 1 of 1339 bp; d,
                     terminator reads only"
     repeat_region   32748. .33022
                     /note "AluJb repeat: matches 1. .289 of consensus"
     repeat_region   33082. .33381
                     /note "15 copies 20 mer 80% conserved"
     repeat_region   33145. .33714
                     /note "38 copies 15 mer 56% conserved"
     repeat_region   33147. .33714
                     /note "284 copies 2 mer or 54% conserved"
     repeat_region   33228. .33499
                     /note "16 copies 17 mer 61% conserved"
     repeat_region   33465. .33724
                     /note "13 copies 20 mer 85% conserved"
     repeat_region   33563. .33710
                     /note "37 copies 4 mer otee 98% conserved"
     repeat_region   33721. .33830
                     /note "MER81 repeat: matches 1. .111 of consensus"
     repeat_region   34241. .34366
                     /note "HERV16 repeat: matches 1698. .1817 of consensus"
     repeat_region   35602. .35673
                     /note "MER5B repeat: matches 1. .67 of consensus"
     repeat_region   35716. .36023
                     /note "AluSx repeat: matches 1. .308 of consensus"
     repeat_region   36696. .36999
                     /note "AluSx repeat: matches 1 .305 of consensus"
     repeat_region   37127. .37234
                     /note "6 copies 18 mer 74% conserved"

alignment scores:
              Quality:     8.00              Length:       8
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_copy_192_200 x AL139123/rev  ..

 Align seq 1/1  to reverse of: AL139123  from: 1  to: 209527

       2 GlnArgSerProLeuGlyGlyAsp 9
         ||||||||||||||||||||||||
  108298 CAGAGGTCCCCTCTGGGAGGTGAT 108275

Seq name: qb ba2:AP003011

Seq documentation block:
LOCUS       AP003011    346510 bp    DNA             BCT       03-FEB-2001
DEFINITION  Mesorhizobium loti DNA, complete genome, section 18/21, complete
            sequence.
ACCESSION   AP003011 BA000012
VERSION     AP003011.1  GI:11994986
KEYWORDS    HTG.
SOURCE      Mesorhizobium loti (strain:MAFF303099) DNA
  ORGANISM  Mesorhizobium loti
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Phyllobacteriaceae; Mesorhizobium.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,T., Nakamura,Y., Sato,S., Asamizu,E., Kato,T., Sasamoto,S.,
            Watanabe,A., Idesawa,K., Ishikawa,A., Kawashima,K., Kimura,T.,
            Kishida,Y., Kiyokawa,C., Kohara,M., Matsumoto,M., Matsuno,A.,
            Mochizuki,Y., Nakayama,S., Nakazaki,N., Shimpo,S., Sugimoto,M.,
            Takeuchi,C., Yamada,M. and Tabata,S.
  TITLE     Complete genome structure of the nitrogen-fixing symbiotic
            bacterium Mesorhizobium loti
  JOURNAL   DNA Res. 7 (6), 331-338 (2000)
  MEDLINE   21082930
REFERENCE   2  (bases 1 to 346510)
  AUTHORS   Kaneko,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
            Institute, The First Laboratory for Plant Gene Research; Yana
            1532-3, Kisarazu, Chiba 292-0812, Japan
            (E-mail:kaneko@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/jp/plant/mlgene/, Tel:81-438-52-3935,
            Fax:81-438-52-3934)
FEATURES             Location/Qualifiers
     source          1. .346510
                     /organism "Mesorhizobium loti"
                     /strain "MAFF303099"
                     /db_xref "taxon:381"
BASE COUNT    62981 a 109633 c 110294 g  63602 t
ORIGIN


alignment scores:
              Quality:     8.00              Length:       8
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_copy_192_200 x AP003011/rev  ..

 Align seq 1/1  to reverse of: AP003011  from: 1  to: 346510

       1 ArgCinArgSerProLeuGlyGly 8
         |||| ||||||||||||||||||||
  199704 CGACAACGATCCCCCCTGGGAGGC 199681

Seq name: qb pat2:AX069957

Seq documentation block:
LOCUS       AX069957         384 bp    DNA             PAT       25-JAN-2001
DEFINITION  Sequence 429 from Patent WO0102568.
ACCESSION   AX069957
VERSION     AX069957.1  GI:12579742
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 384)
  AUTHORS   Williams,L.I., Escobedo,J., Innis,M.A., Garcia,P.D., Klinger,J.,
            Kassam,A., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
            Lamson,G., Drmanac,R., Erkenjakov,R., Drmanac,S., Dickson,M.,
            Labat,I., Leshkowitiz,D., Kita,D., Garcia,V. and Struche Crain,B.
  TITLE     Human genes and gene expression products
  JOURNAL   Patent: WO 0102568-A 429 11-JAN-2001;
            CHIRON CORPORATION (US) ; HYSEQ, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .384
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
     misc_feature    1. .384
                     /note "n   A,T,C or G"
BASE COUNT       65 a    135 c    124 g     59 t       1 others
ORIGIN


alignment scores:
              Quality:     7.00              Length:       7
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_copy_192_200 x AX069957  ..

 Align seq 1/1  to: AX069957  from: 1  to: 384

       1 ArgGlnArgSerProLeuGly 7
         |||| |||||||||||||||
```

```
     137 CGCCAGCGGTCCCCGCTCGGA 157

Seq_name: gb_bat:AF013517

Seq_documentation_block:
LOCUS       AF013517      692 bp    DNA             BCT       23-SEP-1997
DEFINITION  Uncultivated soil bacterium clone C006 16S ribosomal RNA gene,
            partial sequence.
ACCESSION   AF013517
VERSION     AF013517.1  GI:2343189
KEYWORDS    .
SOURCE      uncultivated soil bacterium clone C006.
  ORGANISM  uncultivated soil bacterium clone C006.
            Bacteria; environmental samples; uncultivated soil bacteria clone
            ensemble.
REFERENCE   1  (bases 1 to 692)
  AUTHORS   Kuske,C.R., Barns,S.M. and Busch,J.D.
  TITLE     Diverse uncultivated bacterial groups from soils of the arid
            southwestern United States that are present in many geographic
            regions
  JOURNAL   Appl. Environ. Microbiol. 63 (9), 3614-3621 (1997)
  MEDLINE   97448525
REFERENCE   2  (bases 1 to 692)
  AUTHORS   Barns,S.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-1997) Life Sciences Division, LS-7, Los Alamos
            National Laboratory, M888, Los Alamos, NM 87545, USA
FEATURES             Location/Qualifiers
     source          1..692
                     /organism="uncultivated soil bacterium clone C006"
                     /db_xref="taxon:64202"
                     /note="extracted from Pinyon Pine forest soil, near
                     Cosnino, Arizona"
     rRNA            <1..>692
                     /product="16S ribosomal RNA"
BASE COUNT      165 a    158 c    239 g    129 t      1 others
ORIGIN


Alignment_scores:
            Quality:    7.00            Length:       7
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment_block:
 US-09-126-945B-2_COPY_192_200 x AF013517/rev  ..

 Align seq 1/1  to reverse of: AF013517  from: 1  to: 692

        3 ArgSerProLeuGlyGlyAsp 9
          ||||||||||||||||||||||
      498 CGGTCCCCTCTTGGAGGACAC 478

Seq_name: gb_pat2:AX081564

Seq_documentation_block:
LOCUS       AX081564      1001 bp    DNA             PAT       27-FEB-2001
DEFINITION  Sequence 69 from Patent WO0109350.
ACCESSION   AX081564
VERSION     AX081564.1  GI:13170383
KEYWORDS    .
SOURCE      Haemophilus influenzae.
  ORGANISM  Haemophilus influenzae
            Bacteria; Proteobacteria; Gamma subdivision; Pasteurellaceae;
            Haemophilus.
REFERENCE   1  (bases 1 to 1001)
  AUTHORS   Berthet,F.X., Dalemans,W.L., Denoel,P., Dequesne,G.S., Feron,C.S.,
            Lobet,Y.S., Poolman,J.S., Thiry,G.S., Lhommard,J.S. and Voet,P.S.
  TITLE     Genetically engineered bleb vaccine
  JOURNAL   Patent: WO 0109350-A 69 08-FEB-2001;
            SMITHKLINE BEECHAM BIOLOGICALS S.A. (BE)
FEATURES             Location/Qualifiers
     source          1..1001
                     /organism="Haemophilus influenzae"
                     /db_xref="taxon:727"
BASE COUNT      313 a    196 c    169 g    327 t
ORIGIN


Alignment_scores:
            Quality:    7.00            Length:       7
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment_block:
 US-09-126-945B-2_COPY_192_200 x AX081564/rev  ..

 Align seq 1/1  to reverse of: AX081564  from: 1  to: 1001

        1 ArgGlnArgSerProLeuGly 7
          |||| ||||||||||||||||
      676 AGAcAGAGATCACCTCTGGGT 656

Seq_name: gb_pli:AF037034

Seq_documentation_block:
LOCUS       AF037034      1146 bp    mRNA            PLN       26-NOV-1998
DEFINITION  Zea mays putative peroxidase P7X mRNA, partial cds.
ACCESSION   AF037034
VERSION     AF037034.1  GI:3925232
KEYWORDS    .
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
            Panicoideae; Andropodoneae; Zea.
REFERENCE   1  (bases 1 to 1146)
  AUTHORS   Padeqimas,L.S. and Reichert,N.A.
  TITLE     Isolation of genes and regulatory sequences implicated in
            hypersensitive response from Zea mays nematode-resistant line MI497
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1146)
  AUTHORS   Padeqimas,L.S. and Reichert,N.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-DEC-1997) Department of Plant and Soil Sciences,
            Mississippi State University, Box 9555, Mississippi State, MS
            39762, USA
FEATURES             Location/Qualifiers
     source          1..1146
                     /organism="Zea mays"
                     /db_xref="taxon:4577"
     CDS             <1..>829
                     /codon_start=2
                     /product="putative peroxidase P7X"
                     /protein_id="AAC79954.1"
                     /db_xref="GI:3925233"
                     /translation="sGVNsAVEqLEkVGASLsNLHLHDCYVEGLsASLsNDIssEyS
                     QGPNLILNLPRGFVVVNSIKAQVESVCPGIVSCADILAVAARDsVVSLGGPSWIVLLGR
                     KDSTAAFEGQTSDIIIPTSSLKQLLSATNKKNLNLIDMVALSGAHTIGQAQCSSFNDR
                     IYNESNINCAFAASLKANCPRAGSTALAPLDLITPNAFDNAYYINLLSQKGLLHSDQE
                     LFNNGSTDSTVRSFASQHVGLQQRLKHGHGQDGQPQEVPLDRNPGADQACLTEGQLVNY"
     polyA_site      986
     polyA_site      1011
     polyA_site      1053
     polyA_site      1066
     polyA_site      1146
BASE COUNT      261 a    454 c    308 g    223 t
ORIGIN


Alignment_scores:
            Quality:    7.00            Length:       7
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment_block:
```

```
 US-09-126-945B-2_COPY_192_200 x AF037034/rev    .

 Align seq 1/1  to reverse of: AF037034  from: 1  to: 1146

        1 ArgGlnArgSerProLeuGly 7
          |||||||||||||||||||||
     1176 CGACAACGAAGCCCCTGGGT 1156

seq name: gb_pr9:HSM800884

seq documentation block:
LOCUS        HSM800884     2325 bp     mRNA             PRI       18-FEB-2000
DEFINITION   Homo sapiens mRNA; cDNA DKFZp434J194 (from clone DKFZp434J194).
ACCESSION    AL110227
VERSION      AL110227.1  GI:5817165
KEYWORDS     .
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 2325)
  AUTHORS    Wambutt,R., Reinhner,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
  TITLE      Direct Submission
  JOURNAL    Submitted (15-AUG-1999) MIPS, Am Klopferspitz 18a, D-82152
             Martinsried, GERMANY
COMMENT      Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
             Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
             sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
             consortium of the German Genome Project.
             This clone (DKFZp434J194) is available at the RZPD in Berlin.
             Please contact the RZPD: Ressourcenzentrum, Reinhnerweg 6, 14059
             Berlin Charlottenburg, GERMANY; Email: clone@rzpd.de Further
             information about the clone and the sequencing project is available
             at http://www.sanger.ac.uk/Users/db_pr/cDNA.
FEATURES             Location/Qualifiers
     source          1..2325
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZp434J194"
                     /clone_lib="434 (synonym: hbos4); Vector pSport1; host
                     DH10B; sites NotI + SalI"
                     /dev_stage="adult"
                     /tissue_type="testis"
     polyA_site      2284
BASE COUNT      415 a    748 c    727 g    435 t
ORIGIN


Alignment scores:
            Quality:     7.00            Length:       7
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-126-945B-2_COPY_192_200 x HSM800884    ..

 Align seq 1/1  to: HSM800884  from: 1  to: 2325

        2 GlnArgSerProLeuGlyGly 8
          ||||||||||||||||||||
     1615 CAGCGGAGCCCCTGGGGGGC 1635
```

```
OM of: US-09-126-945B-2_COPY_192_200.rn*  N_Geneseq_0601 *  ** Format  pfs

Date: Sep 14, 2001  9:48 PM

About: Results were produced by the CodonCode software  [***] v 1.5
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL frame+_p2n.model -DEV=xlp
 -Q="/cgn2/USPTO_sp*2/US09126945/Grant_14Sep01_192611_1/US09126945B_Seq_Data_1_1267
 -DB=N_Geneseq_0601 -QFMT=fastap -SUFFIX=olip2n.rng -GAPOP=4.500
 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality
 -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NFSM=ext -MINLEN=9
 -MAXLEN=2000000000 -USER=US09126945_4CON1_1_308 -NCPU=6 -IOPT=3
 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-126-945B-2_COPY_192_200
Query length: 9
Database: N_Geneseq_0601:*
Database sequences: 730101
Database length: 313950809
Search time (sec): 680.150000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000


score_list:
Sequence          Strd Orig    ZScore    EScore Len  ! Documentation  ..
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAA06617  +     9.00  169.21  0.1865     278  !
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAF21928  +     9.00  159.77  0.6284    1087  !
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAZ50951  +     9.00  155.96  1.02      1891  !
/SIDSI/seqdata/geneseq/geneseqn/NA2001.DAT:AAC83261  +     9.00  155.96  1.02      1894  !
/SIDSI/seqdata/geneseq/geneseqn/NA2001.DAT:AAC83266  +     9.00  152.12  1.69      3017  !
/SIDSI/seqdata/geneseq/geneseqn/NA1998.DAT:AAV54975  -     7.00  149.31  2.73        29  !
/SIDSI/seqdata/geneseq/geneseqn/NA2001.DAT:AAF64073  -     7.00  140.60  26.46      384  !
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAC75671  -     7.00  125.40  51.57      825  !
/SIDSI/seqdata/geneseq/geneseqn/NA2001.DAT:AAF91443  -     7.00  124.03  61.45     1001  !
/SIDSI/seqdata/geneseq/geneseqn/NA1998.DAT:AAC57845  -     7.00  120.60  86.42     1451  !
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAF19229  +     7.00  120.47  86.96     1492  !
/SIDSI/seqdata/geneseq/geneseqn/NA1998.DAT:AAV32688  -     7.00  118.01  133.06    2410  !
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAC56015  -     7.00  118.01  133.06    2410  !
/SIDSI/seqdata/geneseq/geneseqn/NA2001.DAT:AAF64019  -     7.00  118.01  133.06    2410  !
/SIDSI/seqdata/geneseq/geneseqn/NA1994.DAT:AAQ66108  -     7.00  117.63  158.91    2919  !
/SIDSI/seqdata/geneseq/geneseqn/NA1994.DAT:AAQ66200  -     7.00  117.63  158.91    2923  !
/SIDSI/seqdata/geneseq/geneseqn/NA1994.DAT:AAQ66201  -     7.00  117.57  160.09    2974  !
/SIDSI/seqdata/geneseq/geneseqn/NA1994.DAT:AAQ66199  -     7.00  116.54  160.56    2984  !
/SIDSI/seqdata/geneseq/geneseqn/NA1994.DAT:AAQ66202  -     7.00  116.54  160.82    2989  !
/SIDSI/seqdata/geneseq/geneseqn/NA2001.DAT:AAF24879  -     7.00  110.43  356.45    7391  !
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAC93213  -     7.00  110.29  359.33    7445  !
/SIDSI/seqdata/geneseq/geneseqn/NA1998.DAT:AAV52953  -     7.00   61.17    1.0e+06  19524  !
/SIDSI/seqdata/geneseq/geneseqn/NA1998.DAT:AAV34379  +     6.00  127.46   39.58      43  !
/SIDSI/seqdata/geneseq/geneseqn/NA1999.DAT:AAX56135  +     6.00  127.46   39.58      43  !
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAC58001  +     6.00  127.46   39.58      43  !
/SIDSI/seqdata/geneseq/geneseqn/NA2001.DAT:AAF60120  +     6.00  127.46   39.58      43  !
/SIDSI/seqdata/geneseq/geneseqn/NA2001.DAT:AAF60176  -     6.00  127.46   39.58      43  !
/SIDSI/seqdata/geneseq/geneseqn/NA2001.DAT:AAF60232  +     6.00  127.46   39.58      43  !
/SIDSI/seqdata/geneseq/geneseqn/NA1996.DAT:AAI14457  +     6.00  124.21   68.32      80  !
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAC12952  +     6.00  119.62  108.24     145  !
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAC12952  +     6.00  117.41  145.50     189  !
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAC20976  +     6.00  115.22  186.00     256  !
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAA41625  +     6.00  114.87  199.10     270  !
/SIDSI/seqdata/geneseq/geneseqn/NA2001.DAT:AAF28002  +     6.00  114.77  201.60     274  !
/SIDSI/seqdata/geneseq/geneseqn/NA2001.DAT:AAF28041  -     6.00  114.77  201.60     274  !
/SIDSI/seqdata/geneseq/geneseqn/NA1996.DAT:AAX36675  +     6.00  114.14  218.43     300  !
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAA01497  +     6.00  114.11  219.42     302  !
/SIDSI/seqdata/geneseq/geneseqn/NA1995.DAT:AAT22209  +     6.00  113.00  225.45     311  !
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAC10432  +     6.00  113.77  229.27     317  !
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAC09179  +     6.00  113.15  248.25     347  !
/SIDSI/seqdata/geneseq/geneseqn/NA1999.DAT:AAY97627  +     6.00  112.88  257.03     361  !
/SIDSI/seqdata/geneseq/geneseqn/NA1999.DAT:AAX41228  +     6.00  112.71  262.66     370  !
/SIDSI/seqdata/geneseq/geneseqn/NA1999.DAT:AAX56151  +     6.00  112.43  271.84     382  !
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAI27799  +     6.00  112.07  285.00     400  !
/SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAC53560  +     6.00  111.38  311.38     449  !

Sequence: /SIDSI/seqdata/geneseq/geneseqn/NA2000.DAT:AAA06617

seq_documentation_block:
ID   AAA06617 standard; cDNA; 278 BP.
XX
AC   AAA06617;
XX
DT   13-JUN-2000  (first entry)
XX
DE   Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:398.
XX
KW   Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW   immunogenic, cytostatic, vaccine, ss.
XX
OS   Homo sapiens.
XX
PN   WO200004149-A2.
XX
PD   27-JAN-2000.
XX
PF   14-JUL-1999;   99WO-US15838.
XX
PR   14-JUL-1998;   98US-0115453.
PR   14-JUL-1998;   98US-0116134.
PR   23-SEP-1998;   98US-0159812.
PR   23-SEP-1998;   98US-0159822.
PR   15-JAN-1999;   99US-0232149.
PR   15-JAN-1999;   99US-0272880.
PR   09-APR-1999;   99US-0288946.
XX
PA   (GENE ) GENXA CORP.
XX
PI   Dillon PP, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;
XX
DR   WPI; 2000-171268/15.
XX
PT   New polypeptide useful for treating and diagnosing prostate cancer
PT   comprises an immunogenic portion of prostate tumor protein -
XX
PS   Claim 50; Page 239; 263pp; English.
XX
CC   The present invention describes isolated polypeptides, comprising an
CC   immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC   and polynucleotides encoding them have cytostatic activity and can be
CC   used in vaccines and in gene therapy. The polypeptides and
CC   polynucleotides encoding them, antigen presenting cells which express
CC   the polypeptides, antibodies against the polypeptides and vaccines
CC   comprising them can be used for inhibiting the development of prostate
CC   cancer in a patient. The polypeptides can be used to generate antibodies
CC   or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC   the polynucleotides encoding the polypeptides can be used as a probe or
CC   to modulate the expression of the polypeptides. AAA06211 to AAA06691 and
CC   AAY82000 to AAY82020 represent sequences used in the exemplification of
CC   the present invention.
XX
SQ   Sequence 278 BP; 56 A; 95 C; 87 G; 39 T; 1 other;


alignment_scores:
            Quality:    9.00              Length:      9
              Ratio:   1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_192_200 x AAA06617  ..
```

```
 Align seq 1/1  to: AAA06617  from: 1  to: 278

        1 ArgGluArgSerProLeuGlyGlyAsp 9
          ||||||||||||||||||||||||||||
       23 CGCCAGCGCTCGCCCCTGGGTGGGGAT 49

seq name: /tmp/pedata/pctoseq_pep_seqs_DNA2* old: /AA1.AA1_28

seq documentation block:
ID   AAF21828 standard; DNA; 1087 BP.
XX
AC   AAF21828;
XX
DT   27-MAR-2001  (first entry)
XX
DE   Human breast and ovarian cancer associated antigen gene SEQ ID 215.
XX
KW   Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
KW   nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW   antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
KW   antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
KW   Addison's disease; allergy; autoimmune haemolytic anaemia;
KW   autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW   multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW   cardiovascular disorder; wound healing; neurological disease; ds.
XX
OS   Homo sapiens.
XX
PN   WO200055177-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05881.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA ) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-611515/58.
DR   P-PSDB; AAB58925.
XX
PT   New human breast and ovarian cancer associated gene sequences and the
PT   polypeptides encoded by these genes, useful in the prevention,
PT   treatment and diagnosis of cancer, immune disorders, cardiovascular
PT   disorders and neurological diseases
XX
PS   Claim 1; Page 642-643; 1299pp; English.
XX
CC   Sequences AAF21614 - AAF22041 represent DNA sequences encoding human
CC   proteins AAB58711 - AAB59128. The DNA and protein sequences are
CC   associated with breast and ovarian cancer. Included in the invention are
CC   sequences AAF22042 - AAF22040 and AAB59129 which are used in the
CC   isolation and characterisation of the DNA and protein sequences of the
CC   invention. The breast and ovarian cancer associated DNA, protein, agonist
CC   or antagonist sequences exhibit cytostatic; immunosuppressive;
CC   nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
CC   antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
CC   antibacterial; antifungal; antiparasitic and cardiant activity. The
CC   polynucleotide and protein sequences are used in the diagnosis of cancer,
CC   particularly breast and ovarian cancer. The nucleic acid sequences,
CC   proteins, agonists and agonists may also be used in the diagnosis,
CC   prevention and treatment of immune disorders e.g. Addison's disease,
CC   allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC   diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC   arthritis and ulcerative colitis; cardiovascular disorders such as
CC   myocardial ischaemias; wound healing; neurological diseases such as
CC   cerebral anoxia and epilepsy, and infectious diseases.
XX
SQ   Sequence 1087 BP; 215 A; 447 C; 339 G; 182 T; 4 other;

alignment scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 Q: ** ** ** ** * * * ** * * * * AAF21828   ..

 Align seq 1/1  to: AAF21828  from: 1  to: 1087

        1 ArgGluArgSerProLeuGlyGlyAsp 9
          ||||||||||||||||||||||||||||
      450 CGCCAGCGCTCGCCCCTGGGTGGGGAT 476

seq name: /tmp/pedata/pctoseq_pep_seqs_DNA2500.1A1.AAZ50691

seq documentation block:
ID   AAZ50691 standard; cDNA; 1894 BP.
XX
AC   AAZ50691;
XX
DT   23-MAY-2000  (first entry)
XX
DE   Human prostate derived Ets factor cDNA.
XX
KW   Prostate derived Ets factor; PDEF; human; chromosome 6p21.3; cancer;
KW   loss of heterozygosity; chromosomal translocation; linkage analysis;
KW   cytostatic; cardiant; immunosuppressive; cerebroprotective; fungicide;
KW   antibacterial; vulnerary; neuroprotective; antiparkinsonian; nootropic;
KW   anabolic; antiinflammatory; anorectic; hybridisation probe; forensic;
KW   tumour marker; diagnosis; treatment; prostate cancer; blood coagulation;
KW   autoimmune disorder; haematopoietic; immune/nervous system stroke;
KW   neoplasm; microbial infection; tissue regeneration; heart attack;
KW   scarring; food additive; preservative; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             416..1423
FT                   /*tag= a
FT                   /product= "Human prostate derived Ets factor"
FT                   /note= "Expressed in prostate epithelium"
XX
PN   WO200006589-A1.
XX
PD   10-FEB-2000.
XX
PF   02-AUG-1999;   99WO-US17470.
XX
PR   31-JUL-1998;   98US-0126945.
XX
PA   (HUMA ) HUMAN GENOME SCI INC.
PA   (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI   Libermann IA,  Oettgen JP,  Kunsch CA,  Endress GA,  Rosen CA;
XX
DR   WPI; 2000-195255/17.
DR   P-PSDB; AAY45003.
XX
PT   Novel p53 state  derived polypeptide, polynucleotide useful for
PT   diagnosis, prevention and treatment of prostate cancer, autoimmune
PT   disorders, microbial infections and also as food additive or
PT   preservative
XX
PS   Claim 4; Fig 1; 132pp; English.
XX
CC   The present cDNA sequence encodes a human prostate-derived Ets factor
CC   (PDEF), isolated from human prostate carcinoma cDNA library
CC   (ATCC No.203072). PDEF gene expression is highest in tissues involved in
CC   androgen sensitivity, with enriched expression in prostate epithelium
CC   cells. PDEF gene is mapped to the human chromosome 6p21.3
CC   region that is associated with loss of heterozygosity and chromosomal
CC   translocations in various human cancers. PDEF has cytostatic, cardiant,
```

CC immunosuppressive, cerebroprotective, fungicide, antibacterial,
CC vulnerary, neuroprotective, antiparkinsonian, nootropic, anabolic,
CC antiinflammatory and anorectic activity. PDEF polynucleotides are useful
CC in linkage analysis as markers, as hybridisation probes for differential
CC identification of the tissues or cell types and as polymorphic markers
CC for forensic purposes. PDEF is useful as prostate-specific tumour marker
CC for the diagnosis and treatment of prostate cancer. PDEF sequences are
CC useful for treating autoimmune disorders, haematopoietic, blood
CC coagulation, immune and nervous system disorders, hyperproliferative
CC disorders like, neoplasms and microbial infections, heart attacks,
CC stroke, scarring and for tissue regeneration. They are also useful as
CC food additives or preservatives.
XX
SQ Sequence 1894 BP; 368 A; 653 C; 571 G; 302 T; 0 other;

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_192_200 x AAZ50691 ..

Align seq 1/1 to: AAZ50691 from: 1 to: 1894

1 ArgGlnArgSerProLeuGlyGlyAsp 9
|||||||||||||||||||||||||||
989 CGCCAGCGCTCGCCCCTGGGTGGGGAT 1015

Seq name: "Sequence 1 from Patent WO0070092"; type: nuc; seq/NA2001:DAI:AAZ50691

seq_documentation_block:
ID AAZ50691 standard; DNA; 1894 BP.
XX
AC AAZ50691;
XX
DT 16-MAR-2001 (first entry)
XX
DE Gland-specific Ets transcription factor (GSEF) nucleotide sequence.
XX
KW Transcription factor; gland specific Ets transcription factor; GSEF;
KW metastatic potential; cancer; tumour; metastasis; breast; prostate;
KW leukaemia; lymphoma; sarcoma; melanoma; chromosome 6p21.1-6p21.3; ds.
XX
OS Homo Sapiens.
XX
PN WO200070092-A1.
XX
PD 23-NOV-2000.
XX
PF 12-MAY 2000; 2000WO-US13173.
XX
PR 14-MAY-1999; 99US-0134112.
XX
PA (CHIR) CHIRON CORP.
XX
PI Kaufmann J, Xin H, Harrowe G;
XX
DR WPI; 2001-041019/05.
DR P-PSDB; AAB49628.
XX
PT Detecting metastatic and potential metastatic cancerous cells, useful
PT for diagnosing, prognosing, grading and staging of cancers by detecting
PT gland-specific Ets transcription factor gene product in a biological
PT sample from a cell -
XX
PS Disclosure; Page 81-83; 95pp; English.
XX
CC This invention relates to a method for the detection and determination of
CC the metastatic potential of a cell. The method comprises detecting a
CC gland-specific Ets transcription factor (GSEF) gene product in a test
CC sample. Detection of a GSEF gene product in the test sample in amount
CC lower than that in a normal cell, is indicative of a cell with high
CC metastatic potential. The method is useful for determining the metastatic
CC potential of a cell, for the diagnosis and prognosis of cancer as well as
CC grading and staging of cancers by detecting GSEF expression in a
CC biological test sample. The method may also be used to monitor patients
CC having a predisposition to develop a particular cancer. GSEF polypeptides
CC are useful for producing antibodies, in cancer diagnosis, prognosis,
CC grading, staging and management of breast and prostate tumours, and in
CC detection polymorphisms in the sequence. GSEF genes and proteins are also
CC useful in gene therapy. GSEF gene product expression levels can be used
CC in conjunction with any tissue in which an alteration in GSEF gene
CC product expression levels is associated with development of a
CC cancer-associated phenotype. Cancers, which can be monitored include
CC cancers of the prostate, cervix, lung and colon, melanomas, colorectal
CC adenocarcinomas, Wilm's tumour, retinoblastoma, sarcomas, myosarcomas,
CC lung carcinomas, leukaemia, and lymphomas. The GSEF gene is located on
CC human chromosome 6, specifically at 6p21.1-6p21.3. The present sequence
CC represents the DNA encoding GSEF.
XX
SQ Sequence 1894 BP; 368 A; 653 C; 571 G; 302 T; 0 other;

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_192_200 x AAC83261 ..

Align seq 1/1 to: AAC83261 from: 1 to: 1894

1 ArgGlnArgSerProLeuGlyGlyAsp 9
|||||||||||||||||||||||||||
989 CGCCAGCGCTCGCCCCTGGGTGGGGAT 1015

Seq name: "Sequence 1 from Patent WO0070092"; type: nuc; seq/NA2001:DAI:AAC83260

seq_documentation_block:
ID AAC83260 standard; DNA; 1917 BP.
XX
AC AAC83260;
XX
DT 16-MAR-2001 (first entry)
XX
DE Gland specific Ets transcription factor (GSEF) cDNA sequence.
XX
KW Transcription factor; gland specific Ets transcription factor; GSEF;
KW metastatic potential; cancer; tumour; metastasis; breast; prostate;
KW leukaemia; lymphoma; sarcoma; melanoma; chromosome 6p21.1-6p21.3; ss.
XX
OS Homo Sapiens.
XX
PN WO200070092-A1.
XX
PD 23-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US13173.
XX
PR 14-MAY-1999; 99US-0134112.
XX
PA (CHIR) CHIRON CORP.
XX
PI Kaufmann J, Xin H, Harrowe G;
XX
DR WPI; 2001-041019/05.
DR P-PSDB; AAB49633.
XX
PT Detecting metastatic and potential metastatic cancerous cells, useful
PT for diagnosing, prognosing, grading and staging of cancers by detecting
PT gland specific Ets transcription factor gene product in a biological
PT sample from a cell -
XX

PS Disclosure; Fig 2; 95pp; English.
XX
CC This invention relates to a method for the detection and determination of
CC the metastatic potential of a cell. The method comprises detecting a
CC gland specific Ets transcription factor (GSEF) gene product in a test
CC sample. Detection of a GSEF gene product in the test sample in amount
CC lower than that in a normal cell, is indicative of a cell with high
CC metastatic potential. The method is useful for determining the metastatic
CC potential of a cell, for the diagnosis and prognosis of cancer as well as
CC grading and staging of cancers by detecting GSEF expression in a
CC biological test sample. The method may also be used to monitor patients
CC having a predisposition to develop a particular cancer. GSEF polypeptides
CC are useful for producing antibodies, in cancer diagnosis, prognosis,
CC grading, staging and management of breast and prostate tumours, and in
CC detecting polymorphisms in the sequence. GSEF genes and proteins are also
CC useful in gene therapy. GSEF gene product expression levels can be used
CC in conjunction with any tissue in which an alteration in GSEF gene
CC product expression levels is associated with development of a
CC cancer associated phenotype. Cancers, which can be monitored include
CC cancers of the prostate, cervix, lung and colon, melanomas, colorectal
CC adenocarcinomas, Wilm's tumour, retinoblastoma, sarcomas, myosarcomas,
CC lung carcinomas, leukaemia, and lymphomas. The GSEF gene is located on
CC human chromosome 6, specifically at 6p21.1 6p21.3. The present sequence
CC represents the cDNA encoding GSEF.
XX
SQ Sequence 3317 BP; 710 A; 1026 C; 970 G; 611 T; 0 other;

alignment scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2 COPY 192_200 x AAC83266 ..

Align seq 1/1 to: AAC83266 from: 1 to: 3317

1 ArgGlnArgSerProLeuGlyGlyAsp 9
||||||||||||||||||||||||||||
2388 CGCCAGCGCTCGCCCCTGGGTGGAGAT 2414

seq name: /SIDS1/gcgdata/geneseq/geneseqp/NA1998.DAT:AAV54975

seq documentation block:
ID AAV54975 standard; DNA; 29 BP.
XX
AC AAV54975;
XX
DT 02-DEC-1998 (first entry)
XX
DE oligonucleotide of the invention.
XX
KW Lambda phage; characterisation; nucleic acid duplex; analyse; detect; ss.
XX
OS Synthetic.
XX
PN WO9837244-A1.
XX
PD 27-AUG-1998.
XX
PF 20-FEB-1998; 98WO-FR000340.
XX
PR 21-FEB-1997; 97FR-0002092.
XX
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Bockelmann U, Beslot F, Essevaz-Roulet B;
XX
DR WPI; 1998-479371/42.
XX
PT Characterising nucleic acid duplex from force required to separate,
PT or re-anneal, strands - used to detect presence of particular
PT structures such as mismatches, translocations, homologous repeats
PT etc.
XX
PS Example 1; Page 17; 62pp; French.
XX
CC oligonucleotides AAV54975-78 were ligated, in various combinations, to
CC DNA2 (a lambda phage sequence serving as a spacer) and to DNA1 (also
CC from lambda phage, the duplex to be opened), and used to exemplify the
CC invention. The specification describes the characterisation of a
CC nucleic acid duplex consisting of at least two partly paired sequences.
CC The method comprises recording at least one signature related to the
CC variation in force needed to separate or reanneal the two strands,
CC then comparing the signature with references. The method is used to
CC analyse unknown, or incompletely known, DNA, and to detect
CC presence/absence of at least one sequence in a sample, particularly to
CC detect a mismatch or for comparison of two DNA samples. Typical features
CC that can be detected are homology, translocations, deletions,
CC amplification, homology repeats, partially paired regions, specific
CC sequence motifs, mutations, CpG islands etc.
XX
SQ Sequence 29 BP; 7 A; 8 C; 10 G; 4 T; 0 other;

alignment scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2 COPY 192_200 x AAV54975/rev ..

Align seq 1/1 to reverse of: AAV54975 from: 1 to: 29

8 ArgSerProLeuGlyGlyAsp 9
|||||||||||||||||||||||
23 CGTAGTCCTTGGGCGGCGAC 4

seq name: /SIDS1/gcgdata/geneseq/geneseqp/NA2001.DAT:AAF64673

seq documentation block:
ID AAF64673 standard; cDNA; 884 BP.
XX
AC AAF64673;
XX
DT 09-AUG-2001 (first entry)
XX
DE Novel human polynucleotide, SEQ ID NO: 429.
XX
KW Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW breast cancer; lung cancer; cancer detection; ss.
XX
OS Homo sapiens.
XX
PN WO200102568-A2.
XX
PD 11-JAN-2001.
XX
PF 30-JUN-2000; 2000WO-US18374.
XX
PR 02-JUL-1999; 99US-0142310.
PR 02-JUL-1999; 99US-0142311.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;
PI Ozkenjikov R, Drmanac S, Dickson M, Lubat L, Leshkowitz D;
PI Kita D, Garcia V, Jones LW, Struche Crain B;
XX
DR WPI; 2001-091808/10.
XX
PT Library of polynucleotides for diagnosing a cancerous state of a

PT mammalian cell and detecting cancer, particularly of the colon or
PT prostate, comprises 3351 human polynucleotide sequences
XX
PS Claim 9; Page 606; 1046pp; English.
XX
CC The present sequence is one of 3351 sequences in a library of human
CC polynucleotides. The library is used to detect differentially expressed
CC genes correlated with a cancerous state of a mammalian cell and can
CC detect colon, prostate, breast and lung cancer. The library can be used
CC to produce probes for detection of mRNA and to produce additional copies
CC of the polynucleotides. The probes can be used for chromosome mapping of
CC the polynucleotide and for detection of transcription levels. Ribozymes
CC or antisense oligonucleotides can be generated. The polynucleotides and
CC their gene products are used as genetic or biochemical markers (e.g. in
CC blood or tissues) that will detect the earliest changes along the
CC carcinogenesis pathway and/or monitor the efficacy of therapies and
CC preventive interventions. The polynucleotides, polypeptides and
CC antibodies against them can be used in pharmaceutical compositions to
CC treat the cancers and proliferative disorders such as neoplasia,
CC dysplasia and hyperplasia
XX
SQ Sequence 884 BP; 65 A; 145 C; 124 G; 59 T; 1 other;

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_192_200 x AAF64673 ..

Align seq 1/1 to AAF64673 from: 1 to: 884

1 ArgGlnArgSerProLeuGly 7
|||||||||||||||||||||
137 CGCCAGCGGTCCCCGCTCGGA 157

seq name: /SIDS1/gcgdata/pepSeq/biosep_NS_... 1&5 AAC75971

seq_documentation_block:
ID AAC75971 standard; cDNA; 820 BP.
XX
AC AAC75971;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF1526 polynucleotide sequence SEQ ID NO:7051.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft versus host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540764.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Lennon M;
XX
DR WPI; 2000-602362/57.
DR P-PSDB; AAB41762.
XX
PT Novel nucleic acids and peptides derived from open reading frames X,
PT useful for treating oral cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease
XX
PS Claim 5; Page 2273-2274; 5507pp; English.
XX
CC AAC74446 to AAC77606 encode the proteins given in AAB40257 to AAB43897,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 820 BP; 192 A; 294 C; 200 G; 132 T; 2 other;

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_192_200 x AAC75971/rev ..

Align seq 1/1 to reverse of: AAC75971 from: 1 to: 820

2 GlnArgSerProLeuGlyGly 8
|||||||||||||||||||||
55 CAGCGAAGCCCTCTGGGTGGG 35

seq name: /SIDS1/gcgdata/pepSeq/biosep_NAC...1&5 AAF91443

seq_documentation_block:
ID AAF91443 standard; DNA; 1001 BP.
XX
AC AAF91443;
XX
DT 04-MAY-2001 (first entry)
XX
DE Haemophilus influenzae (Hi8d) omp26 gene upstream sequence, SEQ ID:69.
XX
KW Modified Gram-negative bacterium; outer membrane vesicle; blebs; vaccine;
KW genetically modified; protective antigen expression; LPS detoxification;
KW LPS; lipid A; homologous recombination vector; immunisation;
KW immunoprotective; non-toxic; paediatric; ds.
XX
OS Haemophilus influenzae.

XX
PN WO200109350 A2.
XX
PD 08-FEB-2001.
XX
PF 31-JUL-2000; 2000WO-EP07424.
XX
PR 03-AUG-1999; 99GB-0018319.
XX
PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Berthet F.X., Dalemans W.L.J., Denoel P., Desquesne G., Feron C., Lobet Y.;
PI Poolman J., Thiry G., Thonnard J., Voet P.;
XX
DR WPI; 2001-138654/14.
XX
PT New isolated polynucleotide useful for outer membrane vesicle
PT preparation from Gram negative bacterial strain for vaccination of
PT microbial infections
XX
PS Claim 46; Page 95; 128pp; English.
XX
CC The invention relates to a genetically engineered outer membrane vesicle
CC (bleb) preparation from a Gram negative bacterium for use as a vaccine.
CC The blebs of the invention are improved with respect to their
CC immunogenicity and toxicity by the introduction of one or more genetic
CC changes to the chromosome of the bacterium from which the blebs are
CC derived. The changes made include the upregulation of protective antigen
CC expression, the downregulation of immunodominant non protective antigen
CC expression, and genetic changes which result in detoxification of the
CC Lipid A moiety of lipopolysaccharide (LPS). The invention also
CC encompasses modified Gram-negative bacterial strains from which the bleb
CC preparations are made, a vector suitable for performing recombination
CC events (for the generation of the modified bacterial strains),
CC bacterially derived nucleic acid sequences used in such a vector, and an
CC immunoprotective and non-toxic Gram-negative bleb, ghost, or killed whole
CC cell vaccine suitable for paediatric use. The bleb preparation is useful
CC in the manufacture of a medicament for immunising a human host against a
CC disease caused by infection of one or more of the following: Neisseria
CC meningitidis, Neisseria gonorrhoeae, Haemophilus influenza, Moraxella
CC catarrhalis, Pseudomonas aeruginosa, Chlamydia trachomatis, and Chlamydia
CC pneumonia. The invention may also be used to provide immunisation against
CC the influenza virus. Bacterially derived nucleotide sequences of the
CC invention are used in the performance of homologous recombination events
CC up to 1000 bp upstream of a bacterial chromosomal gene in order to either
CC increase or decrease expression of that gene. Immunoprotective and
CC non-toxic Gram-negative bleb, ghost, or killed whole cell vaccines
CC are more immunogenic, less toxic and safer, and are particularly useful
CC for paediatric use. The present sequence represents a specifically
CC claimed Haemophilus influenzae nucleic acid sequence.
XX
SQ Sequence 1001 BP; 313 A; 192 C; 169 G; 327 T; 0 other;

alignment scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2-COPY_192_200 x AAF91443/rev .

Align seq 1/1 to reverse of: AAF91443 from: 1 to: 1001

1 ArgGlnArgSerProLeuGly 7
||||||||||||||||||||
658 AGACAGAGATCACCTCTGGGT 638

Seq name: /cgn2_6/ptodata/processed_pone/pone.pp/NA_2 8 ..AAF91443

Seq documentation block:
ID AAZ57845 standard; cDNA; 1651 BP.
XX

AC AAZ57845;
XX
DT 11-APR-2000 (first entry)
XX
DE Protein regulating gene expression PRGE 7 cDNA clone 1734119.
XX
KW Protein regulating gene expression; PRGE 7; human;
KW cancer; inflammation; anticancer; antitumour; antiinflammatory;
KW gene therapy; diagnosis; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 41..711
FT /*tag= a
XX
PN WO9964596-A2.
XX
PD 16-DEC-1999.
XX
PF 11-JUN-1999; 99WO-US13281.
XX
PR 12-JUN-1998; 98US-0089029.
PR 29-JUL-1998; 98US-0094575.
PR 14-OCT-1998; 98US-0104624.
XX
PA (INCY) INCYTE PHARM INC.
XX
PI Lal P., Yue H., Tang Y.T., Hillman J.L., Bandman O., Corley N.C.;
PI Gneppe E.K., Gorgone G.A., Baughn M.R., Patterson C., Lu D.M.;
XX
DR WPI; 2000-116543/10.
DR P-PSDB; AAY58614.
XX
PT New human polypeptides that regulate gene expression, for treatment,
PT prevention and diagnosis of, e.g. cancer -
XX
PS Claim 9; Page 125-126; 150pp; English.
XX
CC The present sequence is that of Incyte clone 1734119 encoding new
CC human protein regulating gene expression PRGE-7 (see AAY58614). The
CC cDNA was initially isolated from Crohn's disease colon tissue cDNA
CC library COLNNOT22, and the full length sequence was assembled from
CC overlapping clones from a number of libraries. PRGE 7 is expressed
CC in reproductive, nervous and gastrointestinal tissues associated
CC cancer, inflammation and foetal diseases, disorders or conditions.
CC The protein product is characterised as a homeodomain protein.
CC The invention provides PRGE polypeptides (see AAY58608-38) and
CC polynucleotides (see AAZ57839-69), expression vectors, host cells,
CC antibodies, agonists and antagonists. It also provides methods for
CC diagnosing, treating or preventing disorders associated with
CC expression of PRGE. Polynucleotides are also used as sources
CC of probes and primers for diagnosis and monitoring of disease, also
CC for detecting related sequences and in gene mapping.
XX
SQ Sequence 1651 BP; 414 A; 507 C; 512 G; 318 T; 0 other;

alignment scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2-COPY_192_200 x AAZ57845 ..

Align seq 1/1 to: AAZ57845 from: 1 to: 1651

[illegible] 8
|||| [illegible]
774 [illegible] 754

Seq name: /cgn2_6/ptodata/processed_pone/pone.pp/NA2000_DAT:AAF18279

```
Seq_documentation_block:
ID   AAF18279 standard; DNA; 1682 BP.
XX
AC   AAF18279;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Lung cancer associated polynucleotide sequence SEQ ID 298.
XX
KW   Human; lung cancer associated protein; neuroprotective; cytostatic;
KW   cardioactive; immunomodulatory; muscular active; vulnerary;
KW   gastrointestinal; nephrotropic; antiinfective; gynecological;
KW   antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW   proliterative disorder; wound healing; infectious disease; ds.
XX
OS   Homo sapiens.
XX
PN   WO200055180-A2.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05918.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (ROSE/) ROSEN C A.
XX
PI   Ruben SM;
XX
DR   WPI; 2000-587514/55.
DR   P-PSDB; AAB58403.
XX
PT   Lung cancer associated gene sequences, referred to as lung cancer
PT   antigens, useful for treatment, prevention, and diagnosis of disorders
PT   such as lung cancer -
XX
PS   Claim 1; Page 757; 1425pp; English.
XX
CC   Polynucleotide sequences AAF17982 - AAF18121 encode human lung cancer
CC   associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC   associated proteins and polynucleotide sequences, their agonists, and
CC   antagonists may have neuroprotective; cytostatic; cardioactive;
CC   immunomodulatory; muscular active general; vulnerary; gastrointestinal
CC   general; nephrotropic; antiinfective; gynecological; or antibacterial
CC   activity. The invention also includes antibodies specific for the
CC   protein or polynucleotide sequences. The lung cancer associated
CC   polynucleotide sequences may be used for detection of lung cancer,
CC   chromosome identification, as chromosome markers, and for numerous other
CC   diagnostic or research purposes. The proteins may be used to treat
CC   disorders such as neural, immune, muscular, reproductive,
CC   gastrointestinal, pulmonary, cardiovascular, renal, and proliterative
CC   disorders. The proteins may also be used in the treatment of wounds and
CC   infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC   peptide AAB58549 are used in the course of the invention for the
CC   identification and characterisation of the polynucleotide and protein
CC   sequences.
XX
SQ   Sequence 1682 BP; 383 A; 551 C; 430 G; 410 T; 8 other;


alignment_scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US 09-126-945B-2_COPY_192_200 x AAF18279    ..

 Align seq 1/1  to AAF18279  from: 1  to:  1682

       1 ArgGluArgSerProLeuGly 7
```

```
                  |||||||||||||||||||
      387 CGCAGCGGTCCCGCTCGGA 407

Seq name: /cgn2_01/data/pea3_seq/fasta_pea3_DNA2000.TXT; AAV42688

Seq_documentation_block:
ID   AAV42688 standard; cDNA; 2410 BP
XX
AC   AAV42688;
XX
DT   20-OCT-1998  (first entry)
XX
DE   Polyomavirus PEA3 cDNA.
XX
KW   Polyomavirus enhancer activator; PEA3; tumour; suppressor; inhibitor;
KW   transformation; HER 2; neu promoter; metastasis; cancer; ss.
XX
OS   Polyomavirus.
XX
FH   Key             Location/Qualifiers
FT   CDS             126..1793
FT                   /*tag= a
FT                   /product= "PEA3
FT                   /note= "polyomavirus enhancer activator"
XX
PN   WO9830585-A2.
XX
PD   16-JUL-1998.
XX
PF   12-JAN-1998;   98WO-US000880.
XX
PR   10-JAN-1997;   97US-0780835.
XX
PA   (TEXA ) UNIV TEXAS SYSTEM.
XX
PI   Hung M.; Xing X;
XX
DR   WPI; 1998-399061/34.
DR   P-PSDB; AAW49010.
XX
PT   Method for repressing transformation of cells - by contacting cell
PT   with polyoma-virus enhancer activator 3, useful for, e.g. treating
PT   or preventing cancer, tumorigenesis and metastasis
XX
PS   Disclosure; Page 69-70; 83pp; English.
XX
CC   This sequence encodes a polyomavirus enhancer activator, PEA3. This
CC   sequence is used in a method for repressing transformation in a cell
CC   which involves contacting the cell with PEA3 to inhibit a transformed
CC   phenotype. This sequence can also be used in a method to suppress the
CC   growth of a tumour in a mammal comprising introducing to the mammal a
CC   PEA3 encoding nucleic acid where the expression of PEA3 in the mammal
CC   results in a decrease in the growth rate of the tumour. PEA3 can regulate
CC   the HER-2/neu promoter by transcriptional repression and acts as a tumour
CC   repressor. The PEA3 can be used for reducing a transforming,
CC   tumorigenic or metastatic potential of a cell. It can be used in the
CC   prevention and treatment of such transformation-driven events as cancer,
CC   tumorigenesis and metastasis.
XX
SQ   Sequence 2410 BP; 521 A; 744 C; 645 G; 500 T; 0 other;


alignment_scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US 09-126-945B-2_COPY_192_200 x AAV42688/rev    ..

 Align seq 1/1  to reverse of: AAV42688  from: 1  to: 2410

       2 GlnArgSerProLeuGlyGly 8
```

```
                    1 1111111111111111111
     1197 CAGCAGAAGCCCTCTGGGTGGA 1177

Seq name: /cgn2_6/ptodata/prod_seq_qc/seq_qc_NA_0_cAC_AA_5_0_1

Seq documentation block:
ID   AAZ50015 standard; cDNA; 2410 BP.
XX
AC   AAZ50015;
XX
DT   25-APR-2000  (first entry)
XX
DE   Murine polyomavirus enhancer activator 3 cDNA.
XX
KW   Murine polyomavirus enhancer activator 3; PEA3; transformation;
KW   tumorigenic, metastatic, cancer, neu mediated cancer, ovarian cancer,
KW   ras mediated cancer; HER/neu promoter; anti transformation therapy;
KW   anti cancer therapy; cytostatic; ss.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             126..1793
FT                   /*tag= a
FT                   /product= "Murine polyomavirus enhancer activator 3"
XX
PN   WO200004153-A2.
XX
PD   27-JAN-2000.
XX
PF   15-JUL-1999;  99WO-US16142.
XX
PR   15-JUL-1998;  98US-0116049.
XX
PA   (TEXA-) UNIV TEXAS SYSTEM.
XX
PI   Hung M;
XX
DR   WPI; 2000-171269/15.
DR   P-PSDB; AAY44724.
XX
PT   Repression of cell transformation used to suppress tumor growth,
PT   comprises contacting the cell with human polyomavirus enhancer
PT   activator 3
XX
PS   Disclosure; Page 86-87; 92pp; English.
XX
CC   The patent discloses methods for repressing transformation in a cell by
CC   contacting it with polyomavirus enhancer activator 3 (PEA3) to inhibit a
CC   transformed phenotype resulting in reduced tumorigenic or metastatic
CC   potential of a cell. This is used in the treatment of various forms of
CC   cancer, e.g. neu- or ras- mediated cancers. The nucleic acid is
CC   introduced into the mammal through a vector or liposomal complex having
CC   DOTMA, DOPE or DC-Chol. The present cDNA sequence encodes murine PEA3.
CC   Murine PEA3 binds directly to HER/neu promoter and represses
CC   transcription in HER2/neu-overexpressing ovarian cancer cells. This can
CC   be used in combination with anti transformation/anti cancer therapies or
CC   chemotherapeutic agents.
XX
SQ   Sequence 2410 BP; 521 A; 744 C; 645 G; 500 T; 0 other;


Alignment scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-126-945B-2 (1-7) _192_200 : AAZ50015 (1-7)

 Align Seq 1/1  to reverse of: AAZ50015  from: 1  to: 2410

       2 GlnArgSerProLeuGlyGly 8
```

```
                    1 1111111111111111111
     1197 CAGCAGAAGCCCTCTGGGTGGA 1177

Seq name: /cgn2_6/ptodata/prod_seq_qc/seq_qc_NA2001_DAI_AAE54019

Seq documentation block:
ID   AAE54019 standard; cDNA; 2410 BP.
XX
AC   AAE54019;
XX
DT   30-MAR-2001  (first entry)
XX
DE   Mouse PEA-3 protein-encoding cDNA, SEQ ID NO:13.
XX
KW   Age-related gene regulation; gene expression; human protein C; hPC;
KW   5' UTR; 5' untranslated region; age-regulatable expression construct;
KW   PEA-3 element; polyoma virus activator-3; antisense therapy;
KW   gene therapy; thrombosis; cardiovascular disease; diabetes;
KW   Alzheimer's disease; Parkinson's disease; cancer; osteoporosis;
KW   osteoarthritis; dementia; ss.
XX
OS   Mus musculus.
XX
PN   WO200075-7-A2.
XX
PD   14-DEC-2000.
XX
PF   06-JUN-2000; 2000WO-US15728.
XX
PR   09-JUN-1999;  99US-0328925.
XX
PA   (UNMI-) UNIV MICHIGAN.
XX
PI   Kurachi K;  Kurachi S;
XX
DR   WPI; 2001-061798/07.
DR   P-PSDB; AAB60290.
XX
PT   New regulatory elements that control age related gene expression,
PT   useful in gene therapy and for reducing Factor IX expression
XX
PS   Disclosure; Fig 9A; 225pp; English.
XX
CC   The invention relates to nucleic acid sequences which regulate gene
CC   expression in an age-related manner and/or in a liver specific manner.
CC   The invention identifies regions of the human factor IX (hFIX) gene, and
CC   a region of the human protein C (hPC) gene, which are age related
CC   regulatory sequences. The hFIX age-related regulatory sequences are
CC   designated AE5' (AAE54016) and AE3' (AAE54017) and are found in the 5'
CC   UTR (at position 2164-2165 of AAE54018) and 3' UTR (at position
CC   4480-4655 of AAE54018) respectively. These elements act synergistically
CC   to increase hFIX levels over the lifespan of an individual; however, they
CC   can independently exert effects on hFIX mRNA in an age related manner,
CC   with AE5' acting to stabilise hFIX mRNA, and AE3' acting to increase hFIX
CC   mRNA levels, over time. AE5' also directs liver specific expression. The
CC   hPC gene age related regulatory sequence is found in the 5' UTR
CC   (AAF54091), and contains two PEA-3 (polyoma virus activator 3) elements
CC   5'-GAGGAAA-3' and 5'-CAGGAAG-3'. The age related regulatory sequences of
CC   the invention, along with their homologues, variants and fragments, may
CC   be used in the construction of recombinant expression vectors for the
CC   expression of a desired sequence in an age related fashion in a host
CC   cell. Preferred target genes for expression in such age regulatable
CC   expression vectors include those encoding proteins involved in blood
CC   coagulation (e.g., the pro-coagulants factor IX and factor VIII, and the
CC   anti-coagulants protein C and antithrombin III), human
CC   alpha 1-antitrypsin, PEA-3 protein and reporter proteins such as
CC   luciferase. Preferred promoters for use in such age regulatable
CC   expression vectors include the human factor IX promoter, the E7 promoter,
CC   the E6 promoter, and the SV40 promoter. The expression vectors of the
CC   invention may be used in gene therapy to provide age- related and/or
CC   liver specific expression of target genes. Age regulatable constructs may
CC   be used in the treatment of such age-related conditions such as
CC   thrombosis, cardiovascular disease, diabetes, Alzheimer's disease,
```

CC Parkinson's disease, cancer, osteoporosis, osteoarthritis and dementia.
CC Specifically, they may be used to express factor IX antisense mRNA in the
CC treatment of thrombotic conditions associated with the natural
CC age-related rise in factor IX expression. Transgenic cells or animals
CC that contain vectors of the invention are useful as models of these
CC diseases. In screening for potential therapeutic agents and for studying
CC normal processes such as ageing and gene expression. Fragments and
CC homologues of age-related regulatory sequences, are useful as probes to
CC detect, isolate or identify other such sequences in samples. The present
CC sequence represents a nucleic acid sequence which may be incorporated
CC into a vector of the invention.
XX
SQ Sequence 2410 BP; 521 A; 744 C; 645 G; 500 T; 0 other;

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_192_200 x AAF54814/rev ..

Align seq 1/1 to reverse of AAF54814 from 1 to 2410

2 GlnArgSerProLeuGlyGly 8
||||||||||||||||||||
1197 CAGAGAAGCCCTCTGGGTGGA 1177

Seq_name: /SIDS1/gcgdata/genseqn/genseqn/NA1_64_DA/AAQ66198

Seq_documentation_block:
ID AAQ66198 standard; DNA; 2949 BP.
XX
AC AAQ66198;
XX
DT 08-DEC-1994 (first entry)
XX
DE H. influenzae b Ca D15 sequence.
XX
KW Vaccine; passive immunization; vector; antiserum; diagnosis; D15;
KW OMP; outer membrane protein; Hib; ss.
XX
OS Haemophilus influenzae type b Ca strain
XX
FH Key Location/Qualifiers
FT CDS 75..2468
FT /*tag= a
XX
PN WO9412641-A.
XX
PD 09-JUN-1994.
XX
PF 23-NOV-1993; 93WO-CA00501.
XX
PR 23-NOV-1992; 92GB-0024584.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Chong P, Klein M, Loosmore S, Sia DYC, Thomas W;
PI Yang Y;
XX
DR WPI; 1994-209269/24.
DR P-PSDB; AAK53754.
XX
PT Nucleic acid encoding D15 outer membrane protein - esp. of
PT Haemophilus influenzae, and related proteins, vectors, antisera
PT etc. useful in vaccines, for diagnosis and for passive
PT immunisation.
XX
PS Disclosure; Fig. 1A; 161pp; English.
XX
CC Outer membrane protein (OMP) D15 genes were isolated by screening
CC chromosomal libraries of H. influenzae type b (Hib) strains Ca,
CC Eagan and Minn A, and the non typeable (NTHi) strains 3655 and 1AK
CC 12085. Nucleotide sequences were determined for the D15 genes
CC (AAQ66198-202) and the corresponding aa sequences were derived
CC (AAK53754-58). D15 OMP can be produced easily and on a large scale,
CC free of other antigens and lipopolysaccharides, by recombinant DNA
CC methods using the isolated genes.
XX
SQ Sequence 2949 BP; 983 A; 477 C; 612 G; 877 T; 0 other;

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_192_200 x AAQ66198 ..

Align seq 1/1 to: AAQ66198 from: 1 to: 2949

1 ArgGlnArgSerProLeuGly 7
|||||||||||||||||||||
[illegible]

```
RM of: US-09-126-945B-2_COPY_192_200.rni  Tuesday, September 11, 2001 ...

Date: Sep 11, 2001  9:53 PM

About: Results were produced by the Smith-Waterman algorithm
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL frame+ p2r model -DEV xhp
 -Q=/cgn2_1/ftptmp/SPTO_spec_1/USPTO_1/945B/app/21_09_11_21_51_3/5_rni_app_pgm.fasta_1 1267
 -DB=Issued_Patents_NA -QFMT fastap -SUFFIX olip2n.rni
 -GAPOP 4.500 -GAPEXT 0.050 -MINMATCH=0.100 -100PCT 0.000
 -100PEXT 0.000 -QGAPOP 4.500 -QGAPEXT=0.050 -XGAPOP 60.000
 -XGAPEXT 60.000 -FGAPOP=6.000 -FGAPEXT=7.000  YGAPOP 60.000
 -YGAPEXT 60.000 -DELOP=6.000 -DELEXT=7.000 -START 1
 -MATRIX oligo -TRANS human40.cdi -LIST 45 -100ALIGN 200
 -THR_SCORE quality -THR_MIN=1  ALIGN=15 -MODE LOCAL  OUTFMT pir
 -NORM ext -MINLEN 0 -MAXLEN=2000000000
 -USER=US09126945_9CGN1_1_91 -NCPU 6 -ICPU 3  LONG=6  N=_XLEXY
 -WAIT -THREADS 1

Search Information block:
Query: US-09-126-945B-2_COPY_192_200
Query length: 9
Database: Issued_Patents_NA:*
Database sequences: 424599
Database length: 94655562
Search time (sec): 422.720000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000


score_list:
Sequence              Strd Orig    ZScore   EScore Len  ! Documentation ...
/cgn2_6/ptodata/2/iia/5B_COMB seq:US-08-780-845B-1      7.00   110.03   63.95     2410  !
/cgn2_6/ptodata/2/iia/6B_COMB seq:US-09-303-268-1       7.00   110.03   63.95     2410  !
/cgn2_6/ptodata/2/iia/5B_COMB seq:US-09-114-216-1       7.00   110.03   63.95     2410  !
/cgn2_6/ptodata/2/iia/6A_COMB seq:US-08-433-522A-2      7.00   110.03   63.95     2949  !
/cgn2_6/ptodata/2/iia/6A_COMB seq:US-09-145-166-2       7.00   110.03   63.95     2949  !
/cgn2_6/ptodata/2/iia/6A_COMB seq:US-08-433-522A-2      7.00   110.03   63.95     2949  !
/cgn2_6/ptodata/2/iia/6A_COMB seq:US-09-145-166-2       7.00   110.03   63.95     2949  !
/cgn2_6/ptodata/2/iia/6A_COMB seq:US-08-433-522A-2      7.00   110.03   63.95     2974  !
/cgn2_6/ptodata/2/iia/6A_COMB seq:US-09-145-166-2       7.00   110.03   63.95     2974  !
/cgn2_6/ptodata/2/iia/6A_COMB seq:US-08-433-522A-2      7.00   110.03   63.95     2984  !
/cgn2_6/ptodata/2/iia/6A_COMB seq:US-09-145-166-2       7.00   110.03   63.95     2984  !
/cgn2_6/ptodata/2/iia/6A_COMB seq:US-08-433-522A-2      7.00   110.03   63.95     2987  !
/cgn2_6/ptodata/2/iia/6A_COMB seq:US-09-145-166-2       7.00   110.03   63.95     2987  !
/cgn2_6/ptodata/2/iia/6A_COMB seq:US-08-433-522A-2      7.00   110.03   63.95     2989  !
/cgn2_6/ptodata/2/iia/6A_COMB seq:US-09-145-166-2       7.00   110.03   63.95     2989  !
/cgn2_6/ptodata/2/iia/6A_COMB seq:US-09-471-985A-91     6.00   122.08   57.33       80  !
/cgn2_6/ptodata/2/iia/PCTUS_COMB seq:PCT_US05_12491A_91     6.00   122.08   57.33       80  !
/cgn2_6/ptodata/2/iia/5B_COMB seq:US-08-914-854-7       6.00   110.03   168.50     459  !
/cgn2_6/ptodata/2/iia/6A_COMB seq:US-09-218-636-7       6.00   110.03   168.50     459  !
/cgn2_6/ptodata/2/iia/5B_COMB seq:US-08-474-479C-68     6.00   110.03   168.50     604  !
/cgn2_6/ptodata/2/iia/6A_COMB seq:US-09-114-216A-68     6.00   110.03   168.50     604  !
/cgn2_6/ptodata/2/iia/6A_COMB seq:US-08-208-198B-68     6.00   110.03   168.50     604  !
/cgn2_6/ptodata/2/iia/5A_COMB seq:US-08-926-427-1       6.00   105.18   304.79    1221  !
/cgn2_6/ptodata/2/iia/5A_COMB seq:US-08-926-427-3       6.00   105.18   304.79    1221  !
/cgn2_6/ptodata/2/iia/6A_COMB seq:US-09-114-939-1       6.00   105.18   304.79    1221  !
/cgn2_6/ptodata/2/iia/6A_COMB seq:US-09-114-939-2       6.00   105.18   304.79    1221  !
/cgn2_6/ptodata/2/iia/6A_COMB seq:US-09-501-672-5       6.00   102.38   417.59    1985  !
/cgn2_6/ptodata/2/iia/6A_COMB seq:US-09-043-111-5       6.00   102.38   417.59    1985  !
/cgn2_6/ptodata/2/iia/5B_COMB seq:US-08-474-479C-79     6.00   101.82   180.22    2041  !
/cgn2_6/ptodata/2/iia/6A_COMB seq:US-09-114-216A-79     6.00   101.82   180.22    2041  !
/cgn2_6/ptodata/2/iia/6A_COMB seq:US-09-208-198B-79     6.00   101.82   180.22    2041  !
/cgn2_6/ptodata/2/iia/PCTUS_COMB seq:PCT_US04-08114-4     6.00   101.45   191.65    2004  !
/cgn2_6/ptodata/2/iia/PCTUS_COMB seq:PCT_US04-12943A-9     6.00   101.45   191.65    2004  !
/cgn2_6/ptodata/2/iia/5B_COMB seq:US-08-864-801-2       6.00   101.27   592.46    2574  !
/cgn2_6/ptodata/2/iia/6A_COMB seq:US-09-265-109-1       6.00   100.74   598.00    2574  !
/cgn2_6/ptodata/2/iia/5B_COMB seq:US-08-816-105A-2      6.00   100.01   501.29    2582  !
/cgn2_6/ptodata/2/iia/6A_COMB seq:US-07-648-452-11      6.00   99.70   615.54    2702  !
/cgn2_6/ptodata/2/iia/5B_COMB seq:US-08-474-479C-11     6.00   99.70   615.54    2702  !
/cgn2_6/ptodata/2/iia/6A_COMB seq:US-09-146-249A-11     6.00   99.70   615.54    2702  !
/cgn2_6/ptodata/2/iia/6A_COMB seq:US-08-208-188B-11     6.00   99.70   615.54    2702  !
/cgn2_6/ptodata/2/iia/6A_COMB seq:US-09-058-489-86      6.00   99.02   671.84    2985  !
/cgn2_6/ptodata/2/iia/5B_COMB seq:US-08-474-479C-64     6.00   97.62   813.88    3705  !
/cgn2_6/ptodata/2/iia/6A_COMB seq:US-09-146-249A-64     6.00   97.62   813.88    3705  !
/cgn2_6/ptodata/2/iia/6A_COMB seq:US-08-208-188B-64     6.00   97.62   813.88    3705  !

>>1_here ../ptodata/2/iia/5B_COMB seq:US-08-780-845B-1

seq_documentation block:
; Sequence 1, Application US/08780845B
; Patent No. 5922688
;   GENERAL INFORMATION:
;     APPLICANT:  Bunz, Mien-Chie
;     APPLICANT:  Xing, Xiaoming
;     TITLE OF INVENTION:  PEA3 is a Tumor Suppressor
;     NUMBER OF SEQUENCES:  9
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  ARNOLD, WHITE AND DURKEE
;       STREET:  P.O. Box 4433
;       CITY:  Houston
;       STATE:  Texas
;       COUNTRY:  USA
;       ZIP:  77210-4433
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/780,845B
;       FILING DATE:  10-JAN-1997
;       CLASSIFICATION:  514
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Wilson, Mark B.
;       REGISTRATION NUMBER:  37,259
;       REFERENCE/DOCKET NUMBER:  UTSC500
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (713) 418-3000
;       TELEFAX:  (713) 474-7577
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2410 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-780-845B-1


alignment_scores:
            Quality:    7.00              Length:       7
              Ratio:    1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_192_200 x US-08-780-845B-1;rev     .

 Alignment of: US-09-126-945B-2_COPY_192_200 to reverse of: US-08-780-845B-1  from: 1  to: 2410

          Matrix: oligo  Apply 8
                  |||||||||||||||||||
    1167 CACACAAGAGGGTCTGGGTGGA 1177


>>1_here ../ptodata/2/iia/6B_COMB.seq:US-09-303-268-1

; Sequence 1, Application US/09303268
; Patent No. 6172212
;   GENERAL INFORMATION:
;     APPLICANT: Bunz, Mien-Chie
;                Xing, Xiaoming
;     TITLE OF INVENTION: PEA3 is a Tumor Suppressor
```

```
;         NUMBER OF SEQUENCES: 9
;         CORRESPONDENCE ADDRESS:
;              ADDRESSEE: ARNOLD, WHITE AND DURKEE
;              STREET: P.O. Box 4433
;              CITY: Houston
;              STATE: Texas
;              COUNTRY: USA
;              ZIP: 77210 4433
;         COMPUTER READABLE FORM:
;              MEDIUM TYPE: Floppy disk
;              COMPUTER: IBM PC compatible
;              OPERATING SYSTEM: PC DOS/MS DOS
;              SOFTWARE: PatentIn Release #1.0, Version #1.30
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/09/403,268
;              FILING DATE: 30 Apr 1999
;              CLASSIFICATION: ·Unknown·
;         PRIOR APPLICATION DATA:
;              APPLICATION NUMBER: 08/780,845
;              FILING DATE: 10 JAN 1997
;         ATTORNEY/AGENT INFORMATION:
;              NAME: Wilson, Mark B.
;              REGISTRATION NUMBER: 37,259
;              REFERENCE/DOCKET NUMBER: UTSC500
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: (512) 418 3000
;              TELEFAX: (512) 474 7577
;    INFORMATION FOR SEQ ID NO: 1:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 2410 base pairs
;              TYPE: nucleic acid
;              STRANDEDNESS: single
;              TOPOLOGY: linear
;         SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US 09 403 268 1


alignment scores:
              Quality:     7.00              Length:       7
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US 09 126 945B 2 COPY_192_200 x US 09 403 268 1/rev   ..

 Align seq 1/1  to reverse of: US 09 403 268 1  from: 1  to: 2410

        2 GlnArgSerProLeuGlyGly 8
          ||||||||||||||||||||
    1197 CAGAGAAGCCCTCTGGGTGGA 1177

seq name: /cm2_6/ptodata/2/US09_COMB_.seq:US 09 116 049 1

seq documentation_block:
; Sequence 1, Application US/09116049A
; Patent No. 6248451
; GENERAL INFORMATION:
;  APPLICANT: Hung, Mien-Chie
;  TITLE OF INVENTION: HUMAN PEA3 IS A TUMOR SUPPRESSOR FOR CANCER CELLS
;  FILE REFERENCE: UTSC:582
;  CURRENT APPLICATION NUMBER: US/09/116,049A
;  CURRENT FILING DATE: 1998 07 15
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 2410
;   TYPE: DNA
;   ORGANISM: Mus musculus
US 09 116 049 1


alignment scores:
              Quality:     7.00              Length:       7
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US 09 126 945B 2 COPY 192 200 x US 09 116 049 1/rev   ..

 Align seq 1/1  to reverse of: US 09 116 049 1  from: 1  to: 2410

        2 GlnArgSerProLeuGlyGly 8
          ||||||||||||||||||||
    1197 CAGAGAAGCCCTCTGGGTGGA 1177

seq name: /cm2_6/ptodata/2/US09_COMB.seq:US 08 433 522A 1

seq_documentation_block:
; Sequence 1, Application US/08433522A
; Patent No. 5015514
;  GENERAL INFORMATION:
;     APPLICANT:  CHONG, Pete
;     APPLICANT:  THOMAS, Wayne
;     APPLICANT:  YANG, Yan Ping
;     APPLICANT:  LOOSMORE, Sheena
;     APPLICANT:  SIA, Hwo Yuan Charles
;     APPLICANT:  KLEIN, Michel
;     TITLE OF INVENTION:  HAEMOPHILUS OUTER MEMBRANE PROTEIN
;     NUMBER OF SEQUENCES:  55
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Sim & McBurney
;       STREET:  6TH Floor, 330 University Avenue
;       CITY:  Toronto
;       STATE:  Ontario
;       COUNTRY:  Canada
;       ZIP:  M5G 1R7
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/433,522A
;       FILING DATE:  12-SEP 1995
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  STEWART, Michael I
;       REGISTRATION NUMBER:  24,973
;       REFERENCE/DOCKET NUMBER:  1048-434 MIS:jb
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (416) 595-1155
;       TELEFAX:  (416) 595-1163
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2949 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  75..2465
US 08 433-522A 1


alignment_scores:
              Quality:     7.00              Length:       7
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US 09 126 945B 2_COPY_192_200 x US 08-433 522A 1   ..

 Align seq 1/1  to: US 08 433 522A 1  from: 1  to: 2949

        1 ArgGlnArgSerProLeuGly 7
```

```
          |||||||||||||||||||
    1913 AGACAGAGATCACCICTGGGT 1933

Seq_name: /cgn2_6/ptodata/2/iom/6A_COMB.seq:US-09-135-166-1

seq_documentation_block:
; Sequence 1, Application US/09135166
; Patent No. 6083743
;  GENERAL INFORMATION:
;    APPLICANT:  CHONG, Pele
;    APPLICANT:  THOMAS, Wayne
;    APPLICANT:  YANG, Yan Ping
;    APPLICANT:  LOOSMORE, Sheena
;    APPLICANT:  SIA, Dwo Yuan Charles
;    APPLICANT:  KLEIN, Michel
;    TITLE OF INVENTION:  HAEMOPHILUS OUTER MEMBRANE PROTEIN
;    NUMBER OF SEQUENCES:  55
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sim & McBurney
;      STREET:  6TH Floor, 330 University Avenue
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5G 1R7
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/135,166
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/433,522
;      FILING DATE:  12-SEP-1995
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEWART, Michael I
;      REGISTRATION NUMBER:  24,973
;      REFERENCE/DOCKET NUMBER:  1038-829 MIS:jb
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (416) 595-1155
;      TELEFAX:  (416) 595-1163
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2949 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  75..2465
US-09-135-166-1


Alignment_scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment_block:
 US-09-126-945B-2_COPY_192_200 x US-09-135-166-1  .

 Align seq 1/1  to: US-09-135-166-1  from: 1  to: 2949

        1 ArgGlnArgSerProLeuGly 7
          |||||||||||||||||||||
    1913 AGACAGAGATCACCICTGGGT 1933

Seq name: /cgn2_6/ptodata/2/iom/6A_COMB.seq:US-09-135-166-1

seq_documentation_block:
; Sequence 5, Application US/08433522A
; Patent No. 6013514
;  GENERAL INFORMATION:
;    APPLICANT:  CHONG, Pele
;    APPLICANT:  THOMAS, Wayne
;    APPLICANT:  YANG, Yan Ping
;    APPLICANT:  LOOSMORE, Sheena
;    APPLICANT:  SIA, Dwo Yuan Charles
;    APPLICANT:  KLEIN, Michel
;    TITLE OF INVENTION:  HAEMOPHILUS OUTER MEMBRANE PROTEIN
;    NUMBER OF SEQUENCES:  55
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sim & McBurney
;      STREET:  6TH Floor, 330 University Avenue
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5G 1R7
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/433,522A
;      FILING DATE:  12-SEP-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEWART, Michael I
;      REGISTRATION NUMBER:  24,973
;      REFERENCE/DOCKET NUMBER:  1038-434 MIS:jb
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (416) 595-1155
;      TELEFAX:  (416) 595-1163
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2950 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  434..2724
US-08-433-522A-5


Alignment_scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment_block:
 US-09-126-945B-2_COPY_192_200 x US-08-433-522A-5  .

 Align seq 1/1  to: US-08-433-522A-5  from: 1  to: 2950

        1 ArgGlnArgSerProLeuGly 7
          |||||||||||||||||||||
    2172 AGACAGAGATCACCICTGGGT 2192

Seq name: /cgn2_6/ptodata/2/iom/6A_COMB.seq:US-09-135-166-5

seq_documentation_block:
; Sequence 5, Application US/09135166
; Patent No. 6083743
;  GENERAL INFORMATION:
;    APPLICANT:  CHONG, Pele
;    APPLICANT:  THOMAS, Wayne
;    APPLICANT:  YANG, Yan Ping
;    APPLICANT:  LOOSMORE, Sheena
;    APPLICANT:  SIA, Dwo Yuan Charles
```

```
;     APPLICANT:  KLEIN, Michel
;     TITLE OF INVENTION:  HAEMOPHILUS OUTER MEMBRANE PROTEIN
;     NUMBER OF SEQUENCES:  55
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Sim & McBurney
;       STREET:  6TH Floor, 330 University Avenue
;       CITY:  Toronto
;       STATE:  Ontario
;       COUNTRY:  Canada
;       ZIP:  M5G 1R7
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/135,166
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/433,522
;       FILING DATE:  12 SEP 1995
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  STEWART, Michael I
;       REGISTRATION NUMBER:  24,973
;       REFERENCE/DOCKET NUMBER:  1038-829 MIS:jb
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (416) 595-1155
;       TELEFAX:  (416) 595-1163
;   INFORMATION FOR SEQ ID NO:  5:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2950 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  334..2724
US-09-135-166-5


alignment_scores:
              Quality:     7.00              Length:       7
                Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2_COPY_192_200 x US-09-135-166-5  ..

 Align seq 1/1  to: US-09-135-166-5  from: 1  to: 2792

        1 ArgGlnArgSerProLeuGly 7
          |||||||||||||||||||||
     2122 AGACAGAGATCACCTCTGGGT 2192

seq name: /seqn2_6/ptodata/2/iru/6A_COMB_seq:US-08-433-522A-7

seq_documentation_block:
; Sequence 7, Application US/08433522A
; Patent No. 6013514
;   GENERAL INFORMATION:
;     APPLICANT:  CHONG, Pole
;     APPLICANT:  THOMAS, Wayne
;     APPLICANT:  YANG, Yan Ping
;     APPLICANT:  LOOSMORE, Sheena
;     APPLICANT:  SIA, Dwo Yuan Charles
;     APPLICANT:  KLEIN, Michel
;     TITLE OF INVENTION:  HAEMOPHILUS OUTER MEMBRANE PROTEIN
;     NUMBER OF SEQUENCES:  55
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Sim & McBurney
;       STREET:  6TH Floor, 330 University Avenue
;       CITY:  Toronto
;       STATE:  Ontario
;       COUNTRY:  Canada
;       ZIP:  M5G 1R7
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/433,522A
;       FILING DATE:  12-SEP-1995
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  STEWART, Michael I
;       REGISTRATION NUMBER:  24,973
;       REFERENCE/DOCKET NUMBER:  1038-434 MIS:jb
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (416) 595-1155
;       TELEFAX:  (416) 595-1163
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2974 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  386..2761
US-08-433-522A-7


alignment_scores:
              Quality:     7.00              Length:       7
                Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2_COPY_192_200 x US-08-433-522A-7  ..

 Align seq 1/1  to: US-08-433-522A-7  from: 1  to: 2974

        1 ArgGlnArgSerProLeuGly 7
          |||||||||||||||||||||
     2224 AGACAGAGATCACCTCTGGGT 2244

seq name: /seqn2_6/ptodata/2/iru/6A_COMB_seq:US-08-433-522A-7

seq_documentation_block:
; Sequence 7, Application US/09135166
; Patent No. 6083743
;   GENERAL INFORMATION:
;     APPLICANT:  CHONG, Pole
;     APPLICANT:  THOMAS, Wayne
;     APPLICANT:  YANG, Yan Ping
;     APPLICANT:  LOOSMORE, Sheena
;     APPLICANT:  SIA, Dwo Yuan Charles
;     APPLICANT:  KLEIN, Michel
;     TITLE OF INVENTION:  HAEMOPHILUS OUTER MEMBRANE PROTEIN
;     NUMBER OF SEQUENCES:  55
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Sim & McBurney
;       STREET:  6TH Floor, 330 University Avenue
;       CITY:  Toronto
;       STATE:  Ontario
;       COUNTRY:  Canada
;       ZIP:  M5G 1R7
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS DOS
```

```
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/135,166
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/433,522
;       FILING DATE:  12-SEP-1995
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  STEWART, Michael I
;       REGISTRATION NUMBER:  24,973
;       REFERENCE/DOCKET NUMBER:  1038-829 MIS:jb
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (416) 595-1155
;       TELEFAX:  (416) 595-1163
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2974 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  486..2761
US-09-135-166-7


alignment_scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_192_200 x US-09-135-166-7    ..

 Align seq 1/1   to: US-09-135-166-7   from: 1   to: 2974

        1 ArgGlnArgSerProLeuGly 7
          |||||||||||||||||||||
     2224 ACACACAGATCACCTCTGGCT 2244

Seq name: /tmp/pubseq/96A_5MB_seq/US-09-135-166-7.seq

seq_documentation_block:
; Sequence 7, Application US/08433522A
; Patent No. 6013514
;  GENERAL INFORMATION:
;     APPLICANT:  CHONG, Pele
;     APPLICANT:  THOMAS, Wayne
;     APPLICANT:  YANG, Yan Ping
;     APPLICANT:  LOOSMORE, Sheena
;     APPLICANT:  SIA, Dwo Yuan Charles
;     APPLICANT:  KLEIN, Michel
;     TITLE OF INVENTION:  HAEMOPHILUS OUTER MEMBRANE PROTEIN
;     NUMBER OF SEQUENCES:  55
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Sim & McBurney
;       STREET:  6TH Floor, 330 University Avenue
;       CITY:  Toronto
;       STATE:  Ontario
;       COUNTRY:  Canada
;       ZIP:  M5G 1R7
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/433,522A
;       FILING DATE:  12-SEP-1995
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  STEWART, Michael I
;       REGISTRATION NUMBER:  24,973
;       REFERENCE/DOCKET NUMBER:  1038-829 MIS:jb
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (416) 595-1155
;       TELEFAX:  (416) 595-1163
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2984 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  874..2764
US-08-433-522A-4


alignment_scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_192_200 x US-08-433-522A-4    ..

 Align seq 1/1   to: US-08-433-522A-4   from: 1   to: 2984

        1 ArgGlnArgSerProLeuGly 7
          |||||||||||||||||||||
     2212 AGACAGAGATCACCTCTGGGT 2232

Seq name: /tmp/pubseq/96A_5MB_seq/US-09-135-166-4.seq

seq_documentation_block:
; Sequence 4, Application US/08433522A
; Patent No. 6083748
;  GENERAL INFORMATION:
;     APPLICANT:  CHONG, Pele
;     APPLICANT:  THOMAS, Wayne
;     APPLICANT:  YANG, Yan Ping
;     APPLICANT:  LOOSMORE, Sheena
;     APPLICANT:  SIA, Dwo Yuan Charles
;     APPLICANT:  KLEIN, Michel
;     TITLE OF INVENTION:  HAEMOPHILUS OUTER MEMBRANE PROTEIN
;     NUMBER OF SEQUENCES:  55
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Sim & McBurney
;       STREET:  6TH Floor, 330 University Avenue
;       CITY:  Toronto
;       STATE:  Ontario
;       COUNTRY:  Canada
;       ZIP:  M5G 1R7
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/135,166
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/433,522
;       FILING DATE:  12-SEP-1995
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  STEWART, Michael I
;       REGISTRATION NUMBER:  24,973
;       REFERENCE/DOCKET NUMBER:  1038-829 MIS.jb
;     TELECOMMUNICATION INFORMATION:
```

```
;       TELEPHONE:  (416) 595 1155
;       TELEFAX:  (416) 595-1163
;     INFORMATION FOR SEQ ID NO:  4:
;       SEQUENCE CHARACTERISTICS:
;         LENGTH:  2984 base pairs
;         TYPE:  nucleic acid
;         STRANDEDNESS:  single
;         TOPOLOGY:  linear
;       MOLECULE TYPE:  DNA (genomic)
;       FEATURE:
;         NAME/KEY:  CDS
;         LOCATION:  374..2764
US-09-135-166-4


alignment_scores:
              Quality:     7.00              Length:       7
                Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_192_200 x US-09-135-166-4   ..

 Align_seq 1/1  to: US-09-135-166-4  from: 1  to: 2984

        1 ArgGlnArgSerProLeuGly 7
          |||||||||||||||||||||
     2212 AGACAGACATCACCTCTCGCT 2232

seq_name: /cgn2_6/ptodata/3/ina/6A_COMB.seq:US-08-433-522A-55

seq_documentation_block:
; Sequence 55, Application US/08433522A
; Patent No. 6014514
;  GENERAL INFORMATION:
;     APPLICANT:  CHONG, Pele
;     APPLICANT:  THOMAS, Wayne
;     APPLICANT:  YANG, Yan Ping
;     APPLICANT:  LOOSMORE, Sheena
;     APPLICANT:  SIA, Dwo Yuan Charles
;     APPLICANT:  KLEIN, Michel
;     TITLE OF INVENTION:  HAEMOPHILUS OUTER MEMBRANE PROTEIN
;     NUMBER OF SEQUENCES:  55
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Sim & McBurney
;       STREET:  6TH Floor, 330 University Avenue
;       CITY:  Toronto
;       STATE:  Ontario
;       COUNTRY:  Canada
;       ZIP:  M5G 1R7
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/433,522A
;       FILING DATE:  12-SEP-1995
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  STEWART, Michael I
;       REGISTRATION NUMBER:  24,973
;       REFERENCE/DOCKET NUMBER:  1038-434 MIS:jb
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (416) 595-1155
;       TELEFAX:  (416) 595-1163
;   INFORMATION FOR SEQ ID NO:  55:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2987 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
; US-08-433-522A-55


alignment_scores:
              Quality:     7.00              Length:       7
                Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_192_200 x US-08-433-522A-55   ..

 Align_seq 1/1  to: US-08-433-522A-55  from: 1  to: 2987

        1 ArgGlnArgSerProLeuGly 7
          |||||||||||||||||||||
     2227 AGACAGACATCACCTCTCGCT 2247

seq_name: /cgn2_6/ptodata/3/ina/6A_COMB.seq:US-09-135-166-55

seq_documentation_block:
; Sequence 55, Application US/09135166
; Patent No. 6083743
;  GENERAL INFORMATION:
;     APPLICANT:  CHONG, Pele
;     APPLICANT:  THOMAS, Wayne
;     APPLICANT:  YANG, Yan Ping
;     APPLICANT:  LOOSMORE, Sheena
;     APPLICANT:  SIA, Dwo Yuan Charles
;     APPLICANT:  KLEIN, Michel
;     TITLE OF INVENTION:  HAEMOPHILUS OUTER MEMBRANE PROTEIN
;     NUMBER OF SEQUENCES:  55
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Sim & McBurney
;       STREET:  6TH Floor, 330 University Avenue
;       CITY:  Toronto
;       STATE:  Ontario
;       COUNTRY:  Canada
;       ZIP:  M5G 1R7
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/135,166
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/433,522
;       FILING DATE:  12-SEP-1995
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  STEWART, Michael I
;       REGISTRATION NUMBER:  24,973
;       REFERENCE/DOCKET NUMBER:  1038-829 MIS:jb
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (416) 595-1155
;       TELEFAX:  (416) 595-1163
;   INFORMATION FOR SEQ ID NO:  55:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2987 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-09-135-166-55


alignment_scores:
              Quality:     7.00              Length:       7
                Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_192_200 x US-09-135-166-55   ..
```

```
 Align seq 1/1  to: US-09-135-166-55  from: 1  to: 2987

       1 ArgGlnArgSerProLeuGly 7
         |||||||||||||||||||||
    2227 AGACAGAGATCACCTCTGGGT 2247

Seq_name: "us2_rpt-data/2_to_6A_COMP_seq:US-09-135-166-9"

seq_documentation_block:
; Sequence 9, Application US/08433522A
; Patent No. 5013514
;  GENERAL INFORMATION:
;     APPLICANT:  CHONG, Pele
;     APPLICANT:  THOMAS, Wayne
;     APPLICANT:  YANG, Yan Ping
;     APPLICANT:  LOOSMORE, Sheena
;     APPLICANT:  SIA, Dwo Yuan Charles
;     APPLICANT:  KLEIN, Michel
;     TITLE OF INVENTION:  HAEMOPHILUS OUTER MEMBRANE PROTEIN
;     NUMBER OF SEQUENCES:  55
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Sim & McBurney
;       STREET:  6TH Floor, 330 University Avenue
;       CITY:  Toronto
;       STATE:  Ontario
;       COUNTRY:  Canada
;       ZIP:  M5G 1R7
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/433,522A
;       FILING DATE:  12-SEP-1995
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  STEWART, Michael I
;       REGISTRATION NUMBER:  24,973
;       REFERENCE/DOCKET NUMBER:  1838-434 MIS:jb
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (416) 595-1155
;       TELEFAX:  (416) 595-1163
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2989 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  390..2768
US-08-433-522A-9


alignment_scores:
            Quality:     7.00              Length:       7
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_192_200 x US-08-433-522A-9     .

 Align seq 1/1  to: US-08-433-522A-9  from: 1  to: 2989

       1 ArgGlnArgSerProLeuGly 7
         |||||||||||||||||||||
    2228 AGACAGAGATCACCGCTGGGT 2248

Seq_name: "us2_rpt_data/2_to_6A_COMP_seq:US-09-135-166-9"
```

```
seq_documentation_block:
; Sequence 9, Application US/09135166
; Patent No. 6083743
;  GENERAL INFORMATION:
;     APPLICANT:  CHONG, Pele
;     APPLICANT:  THOMAS, Wayne
;     APPLICANT:  YANG, Yan Ping
;     APPLICANT:  LOOSMORE, Sheena
;     APPLICANT:  SIA, Dwo Yuan Charles
;     APPLICANT:  KLEIN, Michel
;     TITLE OF INVENTION:  HAEMOPHILUS OUTER MEMBRANE PROTEIN
;     NUMBER OF SEQUENCES:  55
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Sim & McBurney
;       STREET:  6TH Floor, 330 University Avenue
;       CITY:  Toronto
;       STATE:  Ontario
;       COUNTRY:  Canada
;       ZIP:  M5G 1R7
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/135,166
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/433,522
;       FILING DATE:  12 SEP-1995
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  STEWART, Michael I
;       REGISTRATION NUMBER:  24,973
;       REFERENCE/DOCKET NUMBER:  1838 829 MIS:jb
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (416) 595-1155
;       TELEFAX:  (416) 595-1163
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2989 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  390..2768
US-09-135-166-9


alignment_scores:
            Quality:     7.00              Length:       7
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_192_200 x US-09-135-166-9     .

 Align seq 1/1  to: US-09-135-166-9  from: 1  to: 2989

       1 ArgGlnArgSerProLeuGly 7
         |||||||||||||||||||||
    2228 AGACAGAGATCACCGCTGGGT 2248
```

```
M of: US-09-126-945B-2_COPY_192_200 to: EST:*  -out_format : pts

Date: Sep 14, 2001  7:47 PM

About: Results were produced by the GenForce software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
  -MODEL trame+_p2n.model -DEV=xlp
  -Q=/tmp/ [user] -spec=/gcg42641/.rnaut [1] -s /gcg42641/.rnaut/pipe_spec_tmp_1247
  -DB=EST -QFMT=fastap -SUFFIX=olip2n.rst -GAP=P 4.500
  -GAPEXT=0.050 -MINMATCH=0.100 -loop=1 0.000 -loopEXT 0.000
  -QGAP=P 4.500 -QGAPEXT 0.050 -XGAP=P 60.000 -XGAPEXT 60.000
  -YGAP=P 6.000 -YGAPEXT 7.000 -YGAP=P 60.000 -YGAPEXT 60.000
   DELG=P 6.000 -DELEXT 7.000 -START 1 -MATRIX align
  -TRANS human40.cdi -LISF=45 -DOCALIGN=200 -THR_SCORE quality
  -THR_MIN 1 -ALIGN=15 -MODE=LOCAL -OUTFMT pts  -NORM ext  -MINGEN 0
  -MAXLEN 2000000000  USER=US09126945_dGN1_1_4645  NCPU=6
  -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS 1

Search information block.
Query: US-09-126-945B-2_COPY_192_200
Query length: 9
Database: EST:*
Database sequences: 10228115
Database length: 4314594
Search time (sec): 9087.630000

WARN: XGAP=P and YGAP=P must be equal. Assuming YGAP=P=XGAP=P=6.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=0.000


score_list:
Sequence                Strd Orig    ZScore    EScore Len  ! Documentation ..
qb_est78:BE757615         +     9.00  157.77  11.40     539  ! BE757615 212101 MARC 2B0V Bos t
qb_est97:BG179568         +     9.00  154.59  19.49     991  ! BG179568 602428182T1 NIH_MGC_91
qb_est16:AI147947         -     8.00  141.84  88.06     443  ! AI147947 qu62a10.s1 Soares_test
qb_est23:AI187225         -     7.00  146.65  171.20     76  ! AI187225 tp85a06.x1 NCI_CGAP_Pr
qb_est50:AW653624               7.00  144.81  246.53     115  ! AW653624 102397 MARC 1B0V bos t
qb_est71:BE159641         +     7.00  142.51  291.30     139  ! BE159641 MR4 HI0407 1804A0 015
qb_est52:AW868564         +     7.00  141.14  347.80     170  ! AW868564 MR1-BN0063-040500 002
qb_est93:BF892954         +     7.00  130.05  599.53     199  ! BF892954 RC6 MF0140-091100-011
qb_est57:BB135923         +     7.00  128.51  486.69     249  ! BB135923 BB135923 RIKEN full-le
qb_est13:AA922253         +     7.00  127.93  524.86     271  ! AA922253 cm46a05.s1 Soares_NFL
qb_gss7:AQ094953          +     7.00  127.73  537.96     279  ! AQ094953 RG_492F_A1_A12_ME 492
qb_est74:BF962664         +     7.00  127.63  544.75     283  ! BF962664 RC5 NN0183 151200 011
qb_est66:BB484100         -     7.00  127.46  556.59     290  ! BB484100 BB484100 RIKEN full le
qb_est82:BF023607         -     7.00  127.02  588.58     309  ! BF023607 6 Human Bo 1 melanoma
qb_gss9:AQ083492          +     7.00  126.66  617.09     326  ! AQ083492 RPCI11 2B22 T7 RPCI 11
qb_est72:BE229496         +     7.00  126.39  638.61     339  ! BE229496 98MJ0192 Rice Immature
qb_est92:BF748578         +     7.00  125.17  746.88     405  ! BF748578 RC3 BN0412 021300 011
qb_est39:AV630320         +     7.00  125.02  761.47     414  ! AV630320 AV630320 Chlamydomonas
qb_est83:BF071712         +     7.00  124.90  772.80     421  ! BF071712 st26a08.y1 Gm=c1067 G1
qb_est91:BF724594         +     7.00  124.85  777.64     424  ! BF724594 bx06t10.y1 Human Iris
qb_gss17:AQ850229         -     7.00  124.80  782.19     427  ! AQ850229 LMAJTV1_lm04a05.y1 Lei
qb_est19:AI368494         +     7.00  124.76  787.32     430  ! AI368494 qy09e05.x1 NCI_CGAP_Br
qb_gss17:AQ847816         -     7.00  124.63  800.21     438  ! AQ847816 LMAJEV1_lm41e05.y1 Le
qb_est54:AW996649         +     7.00  124.61  801.81     439  ! AW996649 QV3-BN0046-150400-151
qb_est25:AI804387         -     7.00  124.31  833.89     459  ! AI804387 te62i12.x1 Soares_NhHM
qb_est16:AI154144         -     7.00  124.29  835.49     460  ! AI154144 qd77h02.r1 Soares_mamm
qb_est72:BE231888         -     7.00  124.16  849.86     469  ! BE231888 136748 MARC 1P1G Sus s
qb_est50:AW653676         -     7.00  124.03  864.21     478  ! AW653676 102486 MARC 1B0V Bos t
qb_gss41:AZ729161         -     7.00  123.89  880.10     488  ! AZ729161 RPCI-24 79K18.TV1 RPCI
qb_est84:BE190250         +     7.00  123.86  883.28     490  ! BE190250 236744 MARC 2P1G Sus s
qb_est71:BE145388         +     7.00  123.82  888.04     493  ! BE145388 IL5 HF0198-291099-008
qb_est109:T74724          -     7.00  123.56  918.10     512  ! T74724 yo85t06.r1 Soares infant
qb_est93:BF854129         -     7.00  123.46  929.14     519  ! BF854129 MR2 EN0092 261200 B01
qb_gss6:AQ042686          +     7.00  123.31  948.03     531  ! AQ042686 CIT HSP-232708.1F C11
qb_est52:AW844459         -     7.00  123.19  962.16     540  ! AW844459 RC2-CN0051-150200-012
qb_est18:AI325452         -     7.00  123.17  965.43     542  ! AI325452 mp45a09.x1 Barstead MP
qb_est29:AL538358         -     7.00  123.12  971.57     546  ! AL538358 AL538358 LTI_FL013_FB1
qb_gss34:AZ893015         -     7.00  123.03  982.54     553  ! AZ893015 RPCI-24 208H13.TV RPCI
qb_est46:AW375193         -     7.00  122.69  1.0e+03    581  ! AW375193 QV2-CT0121-301199-003
qb_gss5:FR003053          -     7.00  122.57  1.0e+03    591  ! AL027022 Fugu rubripes GSS s
qb_est21:AI513099         -     7.00  122.38  1.1e+03    608  ! AI513099 OH26754.5prime OH 1
qb_gss5:FR003672          -     7.00  122.26  1.1e+03    619  ! Z92482 Fugu rubripes GSS seque
qb_est91:BE678843         -     7.00  122.15  1.1e+03    629  ! BE678843 60215-153P1 NIH_MGC
qb_est53:AW955176         -     7.00  122.13  1.1e+03    630  ! AW955176 EST367246 MAGE reso
qb_est72:BE275993         -     7.00  122.09  1.1e+03    634  ! BE275993 Soares_NFL_T_GBC_S1

seq_name: qb_est78:BE757615

Sequence Type: mRNA
LOCUS       BE757615             539 bp    mRNA            EST       15 SEP 2000
DEFINITION  212101 MARC 2B0V Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BE757615
VERSION     BE757615.1  GI:10171607
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 539)
  AUTHORS   Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
            Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Lacroix
            ,W.W. and Keele,J.W.
  TITLE     Design and use of four pooled tissue normalized cDNA libraries for
            EST discovery in cattle
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            P.O. Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and  minmatch 12 options.
            PCR Primers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 64  row: G  column: 13
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1..539
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 2B0V"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XhoI; Site_2: XhoI;
                     Library made from pooled tissue from testis, thymus,
                     semitendinosus muscle, longissimus muscle, pancreas,
                     adrenal, and endometrium."
BASE COUNT        98 a    168 c    193 g     80 t
ORIGIN


alignment_scores:
            Quality:    9.00            Length:      9
              Ratio:   1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_192_200 x BE757615  ..

 Align seq 1/1  to: BE757615  from: 1  to: 539

        1 ArgGlnArgSerProLeuGlyGlyAsp 9
          ||| ||||||||||||||||||||||||
      326 CGCCAGCGCTCGCCCCTGGGCGGGGAC 352

seq_name: qb_est97:BG179568

seq_documentation_block:
```

```
LOCUS       BG179568      991 bp    mRNA            EST       06-FEB-2001
DEFINITION  602328182F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4429413 5',
            mRNA sequence.
ACCESSION   BG179568
VERSION     BG179568.1  GI:12686271
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo
REFERENCE   1  (bases 1 to 991)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ILSCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10181  row: d  column: 22
            High quality sequence stop: 655.
FEATURES             Location/Qualifiers
     source          1..991
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4429413"
                     /clone_lib="NIH_MGC_91"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage resistant)"
                     /note="Organ: prostate; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.4 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      190 a    351 c    405 g    145 t
ORIGIN


alignment scores:
            Quality:    9.00            Length:       9
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_192_200 x BG179568   ..

 Align seq 1/1  to: BG179568  from: 1  to: 991

       1 ArgGlnArgSerProLeuGlyGlyAsp 9
         |||||||||||||||||||||||||||
     477 CGCCAGCGCTCGCCCCTGGGTGGGGAT 503

seq_name: gb_est16:AI147947

seq_documentation_block:
LOCUS       AI147947      443 bp    mRNA            EST       29-SEP-1998
DEFINITION  qq62a10.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1839738
            3' similar to contains Alu repetitive element;, mRNA sequence.
ACCESSION   AI147947
VERSION     AI147947.1  GI:3675629
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 443)
  AUTHORS   NCI_CGAP http://www.ncbi.nlm.nih.gov/ncicgap
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer:  40ml3 fwd. ET from Amersham
            High quality sequence stop: 429.
FEATURES             Location/Qualifiers
     source          1..443
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1839738"
                     /clone_lib="Soares_testis_NHT"
                     /sex="male"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech Laboratories
                     , Inc., and primed with  a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      100 a    126 c    107 g    110 t
ORIGIN


alignment scores:
            Quality:    8.00            Length:       8
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_192_200 x AI147947/rev   ..

 Align seq 1/1  to reverse of: AI147947  from: 1  to: 443

       2 GlnArgSerProLeuGlyGlyAsp 9
         |||||||||||||||||||||||||
     364 CAGAGGTCCCCTCTGGGAGGTGAT 340

seq_name: gb_est23:AI687225

seq_documentation_block:
LOCUS       AI687225      76 bp    mRNA            EST       14-DEC-1999
DEFINITION  tp85g06.x1 NCI_CGAP_Ut3 Homo sapiens cDNA clone IMAGE:2206138 3'
            similar to SW:EXTN_TOBAC P13983 EXTENSIN PRECURSOR ;, mRNA
            sequence.
ACCESSION   AI687225
VERSION     AI687225.1  GI:4898519
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 76)
  AUTHORS   NCI_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
```

```
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Trace considered overall poor quality
            Insert Length: 580   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 1.
FEATURES              Location/Qualifiers
     source           1..76
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:2206138"
                      /clone_lib "NCI_CGAP_Ut3"
                      /tissue_type "poorly-differentiated endometrial
                      adenocarcinoma, 2 pooled tumors"
                      /lab_host "DH10B"
                      /note "Vector: pCMV SPORT6; Site_1: SalI;
                      Site_2: NotI; Cloned unidirectionally.  Primer: oligo dT.
                      Average insert size 1.45 kb. Life Technologies catalog #:
                      11541-018"
BASE COUNT        16 a      34 c      15 g      11 t
ORIGIN


alignment_scores:
            Quality:    7.00            Length:      7
              Ratio:   1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_192_200 x AI687225    .

 Align seq 1/1  to: AI687225  from: 1  to: 76

        2 GlnArgSerProLeuGlyGly 8
          ||||||||||||||||||||
       47 CAAAGGTCCCCCTTGGGGGGG 67

Seq_name: qb_est50:AW653624

Seq_documentation_block:
LOCUS         AW653624       115 bp    mRNA            EST       11-JUL-2000
DEFINITION  102497 MARC 1B0V Bos taurus cDNA 5', mRNA sequence.
ACCESSION   AW653624
VERSION     AW653624.1  GI:7419450
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 115)
  AUTHORS   Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
            Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Lacqreid
            ,W.W. and Keele,J.W.
  TITLE     Design and use of four pooled tissue normalized cDNA libraries for
            EST discovery in cattle
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166  Clay Center  NE 68933-166  USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and clt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR Primers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 107  row: B  column: 4
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES              Location/Qualifiers
     source           1..115
                      /organism="Bos taurus"
                      /db_xref="taxon:9913"
                      /clone_lib "MARC 1B0V"
                      /tissue_type "pooled"
                      /lab_host "DH10B"
                      /note "Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                      Library made from pooled tissue from lymph node, ovary,
                      fat, hypothalamus, and pituitary."
BASE COUNT        17 a      41 c      35 g      24 t
ORIGIN


alignment_scores:
            Quality:    7.00            Length:      7
              Ratio:   1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_192_200 x AW653624/rev    .

 Align seq 1/1  to reverse of: AW653624  from: 1  to: 115

        1 ArgLeuArgSerProLeuGly 7
          ||| |||||||||||||||||
       59 CGGCAAAGAAGTCCTCTCCGGT 39

Seq_name: qb_est71:BE159641

Seq_documentation_block:
LOCUS         BE159641       139 bp    mRNA            EST       21-JUN-2000
DEFINITION  MRC-HT0407 1B0430-015-b11 HT0407 Homo sapiens cDNA, mRNA sequence
ACCESSION   BE159641
VERSION     BE159641.1  GI:8622362
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 139)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Bronstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo - SP,
            Brazil
            Tel: +55-11 2704922
            Fax: +55 11 2707061
            Email: asimpson@ludwig.org.br
            This sequence was derived from the PAFGE/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            [illegible] MRC-HT0407 1B0
            430-015 b11&t3 2000-04-18&t4 1)
            Seq primer: puc 18 forward
            High quality sequence start: 9
            High quality sequence stop: 139.
FEATURES              Location/Qualifiers
     source           1..139
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone_lib="HT0407"
                      /dev_stage="Adult"
```

```
                 /note "organ: head neck; Vector: puc18; Site 1: SmaI;
                 Site 2: SmaI; A mini library was made by cloning products
                 derived from  ORESTES PCR (U.S. Letters Patent application
                 No. 196,716    Ludwig Institute for Cancer Research)
                 profiles into  the pUC 18 vector. Reverse transcription of
                 tissue mRNA and cDNA amplification were performed under
                 low stringency conditions."
BASE COUNT       22 a     46 c     39 g     31 t      1 others
ORIGIN


alignment_scores:
           Quality:    7.00              Length:      7
             Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_192_209 x BE159641   ..

 Align seq 1/1  to: BE159641  from: 1  to: 199

        1 ArgGlnArgSerProLeuGly 7
          |||||||||||||||||||||
       97 AGGCAGCGCAGTCCGCTGGGA 117

seq name: gb_est52:AW868564

seq documentation block:
LOCUS       AW868564       170 bp    mRNA             EST        22 MAY 2000
DEFINITION  MR1-SN0063-040500-002-a03 SN0063 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW868564
VERSION     AW868564.1  GI:8002616
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 170)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509 010, Sao Paulo SP,
            Brazil
            Tel: +55 11 2704922
            Fax: +55 11 2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1 &t2 MR1-SN0063-040
            500-002-a03&t3 2000-05-04&t4 1)
            Seq primer: puc 18 forward
            High quality sequence start: 30
            High quality sequence stop: 93.
FEATURES             Location/Qualifiers
     source          1..170
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone_lib "SN0063"
                     /dev_stage "Adult"
                     /note "organ: stomach normal; Vector: puc18; Site 1: SmaI;
                     Site 2: SmaI; A mini library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716    Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       19 a     39 c     66 g     46 t
ORIGIN


alignment scores:
           Quality:    7.00              Length:      7
             Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_192_200 x AW868564   ..

 Align seq 1/1  to: AW868564  from: 1  to: 170

        2 GlnArgSerProLeuGlyGly 8
          |||| ||||||||||||||||
      126 CAGCGTCTCCTCTGGGGGC 146

seq name: gb_est93:BF892954

seq documentation block:
LOCUS       BF892954       199 bp    mRNA             EST        18 JAN 2001
DEFINITION  RC6-MT0140-091100-011-H07 MT0140 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF892954
VERSION     BF892954.1  GI:12284517
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 199)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509 010, Sao Paulo SP,
            Brazil
            Tel: +55 11 2704922
            Fax: +55 11 2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1 &t2 RC6-MT0140
            091100-011-H07&t3 2000-11-09&t4 1)
            Seq primer: puc 18 forward
            High quality sequence start: 19
            High quality sequence stop: 199.
FEATURES             Location/Qualifiers
     source          1..199
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone_lib "MT0140"
                     /dev_stage "Adult"
                     /note "organ: marrow; Vector: puc18; Site 1: SmaI; Site 2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716    Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT       42 a     61 c     49 g     47 t
```

```
ORIGIN


alignment_scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_192_200 x BB842454

 Align seq 1/1  to:  BB842454  from:  1  to:  112

        1 ArgGlnArgSerProLeuGly 7
          |||||||||||||||||||
       92 AGGCAGACATCGCCTCTTGGT 112

seq_name: gb_est57:BB135923

seq_documentation_block:
LOCUS       BB135923       249 bp    mRNA            EST       28-JUN-2000
DEFINITION  BB135923 RIKEN full-length enriched, adult male Mus musculus
            cDNA clone 9830131K20 3', mRNA sequence.
ACCESSION   BB135923
VERSION     BB135923.1  GI:8790860
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 249)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kamiya,T., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,N., Mizuno,Y., Nishimura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,R., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shimagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Taqawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,T.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshito
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp,
            URL:http://genome.rtc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M., Hirai,H., 2000
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
            Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
            Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 464-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            High efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
            Please visit our web site (http://genome.rtc.riken.go.jp/) for
            further details.
FEATURES             Location/Qualifiers
     source          1..249
                     /organism "Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="9830131K20"
                     /clone_lib "RIKEN full-length enriched, adult male 1 of"
                     /sex="male"
                     /tissue_type "bone"
                     /dev_stage "adult"
                     /lab_host "DH10B"
                     /note "Site 1: SalI; Site 2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Gene Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN, Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3']. cDNA was
                     prepared by using trehalose thermoactivated reverse
                     transcriptase and subsequently enriched for full length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot  10.0 and subtraction to Rot  185.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGAAGGATCCGAGCTTAATTAAATTAATCCCCCCCCCCCCCCCC
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(-) after bulk excision from Lambda
                     FLC I."
BASE COUNT        71 a      58 c      48 g      72 t
ORIGIN


alignment_scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_192_200 x BB135923

 Align seq 1/1  to:  BB135923  from:  1  to:  249

        1 ArgGlnArgSerProLeuGly 7
          |||||||||||||||||||
       97 AGACAGATCGCCTCTTGGCA 97

seq_name: gb_est15:AA922253

seq_documentation_block:
LOCUS       AA922253       271 bp    mRNA            EST       23-JUN-1998
DEFINITION  zm46c05.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:1544096 3', mRNA sequence.
ACCESSION   AA922253
VERSION     AA922253.1  GI:3069562
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo
REFERENCE   1  (bases 1 to 271)
  AUTHORS   [illegible]
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 504  Std Error: 0.00
            Seq primer: -40mUL fwd. ET from Amersham
            High quality sequence stop: 254.
FEATURES             Location/Qualifiers
     source          1..271
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone "IMAGE:1544096"
                     /clone_lib "Soares_NFL_T_GBC_S1"
```

```
                     /lab_host "DH10B"
                     /note "Organ: pooled; Vector: pT7T3D Pac (Pharmacia) with
                     a modified polylinker; Site 1: Not I; Site 2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B cell
                     NC1_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682642-687239,
                     726408-728711, and 729096-733899. Subtraction by Bento
                     Soares and M. Fatima Bonaldo. "
BASE COUNT        71 a      56 c      58 g      86 t
ORIGIN


alignment scores:
              Quality:     7.00              Length:       7
                Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2_COPY_192_200 x AA922253     .

 Align Seq 1/1   to: AA922253  from: 1  to: 271

        1 ArgGlnArgSerProLeuGly 7
          |||||||||||||||||||||
       48 CGTCAGAGGTCACCTTTAGGT 58

seq_name: qb qss7:AQ094853

seq documentation block:
LOCUS        AQ094853       279 bp     DNA              GSS        27-AUG-1998
DEFINITION   HS_3029_A1_A12_MR CIT Approved Human Genomic Sperm Library D Homo
             sapiens genomic clone Plate 3029 Col 23 Row A, DNA sequence.
ACCESSION    AQ094853
VERSION      AQ094853.1  GI:3466400
KEYWORDS     GSS.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 279)
  AUTHORS    Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
             Kelfer,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
             Hood,L.
  TITLE      Sequence tagged connectors: A sequence approach to mapping and
             scanning the human genome
  JOURNAL    Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE    99380589
COMMENT      Contact: Mahairas GG; Wallace JC; Hood L
             High Throughput Sequencing Center
             University of Washington
             401 Queen Anne Avenue North, Seattle, WA 98109, USA
             Tel: (206) 616-4618
             Fax: (206) 616-4887
             Email: jwallace@u.washington.edu
             Sequence Tagged Connector
             Plate: 3029  row: A  column: 23
             Class: BAC ends
             High quality sequence stop: 279.
FEATURES             Location/Qualifiers
     source          1..279
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate: 3029 Col 23 Row A"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E coli DH10B"
BASE COUNT        70 a      71 c      66 g      72 t
ORIGIN


alignment scores:
              Quality:     7.00              Length:       7
                Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2_COPY_192_200 x AQ094853     .

 Align Seq 1/1   to: AQ094853  from: 1  to: 279

        2 GlnArgSerProLeuGlyGly 8
          |||||||||||||||||||||
      160 CAGAGGTCACCTTTAGGAGG 180

seq_name: qb est94:BF962664

seq documentation block:
LOCUS        BF962664       283 bp     mRNA             EST        22-JAN-2001
DEFINITION   RC_NN0183_151200_011_E11_NNG183 Homo sapiens cDNA, mRNA sequence
ACCESSION    BF962664
VERSION      BF962664.1  GI:12379939
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 283)
  AUTHORS    Dias Neto,E., Garcia Correa,R., Verjovski Almeida,S., Briones,M.R.,
             Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
             Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
             Brumstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
             ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
             Simpson,A.J.
  TITLE      Shotgun sequencing of the human transcriptome with ORF expressed
             sequence tags
  JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE    20202663
COMMENT      Contact: Simpson A.J.G.
             Laboratory of Cancer Genetics
             Ludwig Institute for Cancer Research
             Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo SP,
             Brazil
             Tel: +55-11-2704922
             Fax: +55-11-2707001
             Email: asimpson@ludwig.org.br
             This sequence was derived from the FAPESP/LICR Human Cancer Genome
             Project. This entry can be seen in the following URL
             http://www.ludwig.org.br/cgi-bin/ace/getmult2.pl?t1=EST&t2=RC_NN0183
             _151200_011_E11&t3=2000-12-15&t4=1)
             Seq primer: puc 18 forward
             High quality sequence stop: 261
FEATURES             Location/Qualifiers
     source          1..283
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="NN0183"
                     /dev_stage="Adult"
                     /note="Organ: nervous normal; Vector: puc18; Site 1: Sma1;
                     Site 2: Sma1; A mini library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716    Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT        70 a      61 c     109 g      63 t
ORIGIN


alignment scores:
              Quality:     7.00              Length:       7
```

```
            Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_192_200 x BF962664  ..

 Align seq 1/1  to:: BF962664  from: 1  to:: 284

       2 ClnArgSerProLeuGlyGly 8
         ||||||||||||||||||||
      20 CAGAGAAGCCCGTTAGGAGGC 40

seq_name: gb_est66:BB484100

seq_documentation_block:
LOCUS       BB484100       290 bp    mRNA            EST       24-JUL-2000
DEFINITION  BB484100 RIKEN full-length enriched, 13 days embryo lung Mus
            musculus cDNA clone D430023K22 3', similar to O11147 Human ataxia
            nucleotide regulatory protein (ABR) mRNA, mRNA sequence.
ACCESSION   BB484100
VERSION     BB484100.1  GI:9401709
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 290)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,I., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,I., Itoh,M.,
            Izawa,M., Kadota,K., Kanjiwa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,I., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,I., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESIs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp,
            URL:http://genome.rtc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
            Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Izawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
            Automated filtration-based high throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1..290
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="D430023K22"
                     /clone_lib="RIKEN full length enriched, 13 days embryo
                     lung"
                     /tissue_type="lung"
                     /dev_stage="13 days embryo"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA
                     was cleaved with BamHI and XhoI. Vector: a modified
                     pBluescript KS(-) after bulk excision from Lambda ZAP II."
BASE COUNT        74 a      98 c      55 g      63 t
ORIGIN


alignment scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_192_200 x BF962664/rev  ..

 Align seq 1/1  to reverse of: BB484100  from: 1  to: 290

       3 ArgSerProLeuGlyGlyAsp 9
         |||||||||||||||||||||
      77 AGGTCCTCCTTTAGGCCGCCAC 57

seq_name: gb_est82:BF023607

seq_documentation_block:
LOCUS       BF023607       309 bp    mRNA            EST       08-NOV-2000
DEFINITION  6 Human HO-1 melanoma cells Homo sapiens cDNA clone mda-D-63, mRNA
            sequence.
ACCESSION   BF023607
VERSION     BF023607.1  GI:11095497
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 309)
  AUTHORS   Jiang,H., Kang,D.-c., Alexandre,D. and Fisher,P.B.
  TITLE     RAID, a rapid subtraction hybridization approach for identifying
            and cloning differentially expressed genes
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (23), 12684-12689 (2000)
COMMENT     Contact: Fisher PB
            Department of Urology
            Columbia University, P & S
            630 W168th St., New York, NY 10032, USA
            Tel: 212 305 3441
            Fax: 212 305 4262
            Email: pbf1@columbia.edu
            Insert Length: 309  Std Error: 0.00
            POLYA No.
FEATURES             Location/Qualifiers
     source          1..309
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="mda-D-63"
                     /clone_lib="Human HO-1 melanoma cells"
                     /cell_type="HO-1 melanoma cells"
BASE COUNT       106 a      77 c      80 g      46 t
ORIGIN
```

```
alignment scores:
            Quality:     7.00              Length:       7
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_192_200 x BF023607/rev    .

 Align seq 1/1  to reverse of: BF023607  from: 1  to: 409

        2 GluArgSerProLeuGlyGly 8
          ||||||||||||||||||||||
      406 CAGCGAAGCCCTCTGGGTGGG 286

Seq name: qb qss9:AQ283492

Seq documentation block:
LOCUS       AQ283492       326 bp    DNA              GSS       27-APR-1999
DEFINITION  RPCI11-2D22.TJ RPCI-11 Homo sapiens genomic clone RPCI-11 2D22, DNA
            sequence.
ACCESSION   AQ283492
VERSION     AQ283492.1  GI:3908409
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 326)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Ross,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wilde,C., de Jong,P. and Venter,J.C.
  TITLE     Use of human BAC End Sequences for Sequence Ready Map Building
  JOURNAL   Unpublished (1998)
COMMENT     Other GSSs: RPCI11-2D22.TV
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.chori.org/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/db/humgen/bac_end/rpci_bac_end_search.html
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..326
                     /organism="Homo sapiens"
                     /db_xref="GDB:7500477"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-2D22"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT       84 a     63 c     87 g     92 t
ORIGIN


alignment scores:
            Quality:     7.00              Length:       7
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_192_200 x AQ283492    ..

 Align seq 1/1  to: AQ283492  from: 1  to: 326

        .
```

```
        3 ArgSerProLeuGlyGlyAsp 9
          |||| ||||||||||||||||||
      277 AGAACCCTCTCGGACGGTGAC 297
```

```
OM of: US-09-126-945B-2_COPY_215_222 to: GenEmbl:*   Out_format : ptx

Date: Sep 14, 2001  9:39 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
  MODEL frame=p2n.model -DEV=xlp
 -Q "/cgn2_2/ptodata/1/ispcpl/US09_126_945B_2_COPY_215_222_1/app_query.fasta_1.1/67
 -DB=GenEmbl -QFMT fastap  SUFFIX olip2n.rge  GAPOP=4.500
 -GAPEXT 0.050 -MINMATCH 0.100 -LOOPT=0.000 -LOOPEXT 0.000
 -QGAPOP 4.500 -QGAPEXT 0.050 -XGAPOP=60.000 -XGAPEXT=60.000
 -FGAPOP=6.000 -FGAPEXT 7.000 -YGAPOP=60.000 -YGAPEXT 60.000
 -DELOP=6.000 -DELEXT 7.000 -START 1 -MATRIX oligo
 -TRANS human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality
  THR_MIN 1  ALIGN 15  MORE [=n]  -PTFMT pir  -NORM=z  -MINLEN=
 -MAXLEN=2000000000 -USER US09126945_COGN1_1_7594 -NCPU 6
  ICPU 3  -LONGLOG -NO_XLPXY -WAIT -THREADS 1

Search information block:
Query: US-09-126-945B-2_COPY_215_222
Query length: 8
Database: GenEmbl:*
Database sequences: 1844157
Database length:  856060004
Search time (sec): 6504.700000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP on user
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000


score_list:
Sequence              Strd Orig    ZScore    EScore Len  ! Documentation ..
qb_rel:AB019436         +     8.00  145.26   7.4e    1704  ! AB019436 Mus musculus mRNA for
qb_pat1:AX048148        +     8.00  144.59   8.1e    1894  ! AX048148 Sequence 1 from Paten
qb_pr4:AF071538         +     8.00  144.59   8.1e    1894  ! AF071538 Homo sapiens Ets tran
qb_pr1:AB031549         +     8.00  144.58   8.1e    1898  ! AB031549 Homo sapiens PSE mRNA
qb_pat1:AX048159        +     8.00  141.08   12.74   3817  ! AX048159 Sequence 12 from Pate
qb_pr6:AL157372         +     8.00  117.16   274.92  155663  ! AL157372 Human DNA sequence
qb_rel:MUSFRD           +     7.00  141.88   11.5e   141   ! L39118 Mus musculus (clone DE 1
qb_vi2:HIV3A9LTR        +     7.00  131.94   41.1e   688   ! L28885 Human immunodeficiency v
qb_vi2:HIV3B7LTR        +     7.00  131.89   41.3e   688   ! L28886 Human immunodeficiency v
qb_ro2:RATCEEIB         +     7.00  127.30   74.5e   1442  ! D00362 Rattus norvegicus mRNA
qb_ba1:AE001735         +     7.00  115.19   352.4e  9949  ! AE001735 Thermotoga maritima s
qb_in3:CELC24H8         +     7.00  106.51   1.1e+03 39721  ! U80453 Caenorhabditis elegans
qb_pr2:AC005756         +     7.00  105.97   1.1e+03 44299  ! AC005756 Homo sapiens chromos
qb_htq17:AC080145       +     7.00  100.09   2.4e+03 110697  ! AC080145 Homo sapiens chromo
qb_htq19:AL139159       -     7.00   99.15   2.8e+03 128490  ! AL139159 Homo sapiens chromo
qb_htq3:AC012233        -     7.00   98.64   2.9e+03 139514  ! AC012233 Homo sapiens clone
qb_htq10:AC026926       -     7.00   98.20   3.1e+03 119388  ! AC026926 Homo sapiens chromo
qb_htq6:AC018847        -     7.00   98.17   3.1e+03 150303  ! AC018847 Homo sapiens chromo
qb_htq25:CNS07ED2       -     7.00   98.08   3.2e+03 152469  ! AC115413 Homo sapiens chromo
qb_htq7:AC020579        +     7.00   98.04   3.2e+03 154596  ! AC020579 Homo sapiens clone
qb_htq11:AC027420       -     7.00   98.04   3.2e+03 154717  ! AC027420 Homo sapiens clone
qb_htq14:AC090669       +     7.00   97.98   3.2e+03 155003  ! AC090669 Homo sapiens chromo
qb_pr6:AL358777         -     7.00   97.96   3.2e+03 155459  ! AL358777 Human DNA sequence
qb_htq23:AL590434       -     7.00   97.76   3.3e+03 160151  ! AC093434 Homo sapiens chromo
qb_htq5:AC016806        +     7.00   97.69   3.4e+03 162112  ! AC016806 Homo sapiens clone
qb_htq16:AC074123       +     7.00   97.50   3.4e+03 167275  ! AC074123 Homo sapiens chromo
qb_htq4:AC013715        -     7.00   97.15   3.6e+03 176705  ! AC013715 Homo sapiens chromo
qb_pr5:AF265340         -     7.00   97.03   3.6e+03 180348  ! AF265340 Homo sapiens chromo
qb_htq22:AL442635       +     7.00   96.98   3.6e+03 181687  ! AL442635 Homo sapiens chromo
qb_htq8:AC023191        -     7.00   96.90   3.7e+03 184109  ! AC023191 Homo sapiens chromo
qb_htq15:AC073118       -     7.00   96.88   3.7e+03 184538  ! AC073118 Homo sapiens chromo
qb_htq8:AC022281        +     7.00   96.82   3.7e+03 186431  ! AC022281 Homo sapiens chromo
qb_htq7:AC021585        +     7.00   96.79   3.7e+03 187418  ! AC021585 Homo sapiens clone
qb_htq4:AC015456        +     7.00   96.76   3.7e+03 188030  ! AC015456 Homo sapiens chromo
qb_htq16:AC073863       -     7.00   96.72   3.8e+03 189287  ! AC073863 Homo sapiens chromo
qb_pr2:AC006994         +     7.00   96.63   3.8e+03 192104  ! AC006994 Homo sapiens BAC cl
qb_htq7:AC021195        +     7.00   96.60   3.8e+03 193007  ! AC021195 Homo sapiens chromo
qb_htq20:AL354803       +     7.00   96.56   3.8e+03 194322  ! AL354803 Homo sapiens chromo
qb_pr7:CNS01RG7         +     7.00   96.54   3.9e+03 194754  ! AL157792 Human chromosome 14
qb_htq24:CNS01DKV       +     7.00   96.54   3.9e+03 194870  ! AL118650 Homo sapiens chr
qb_htq2:AC009993        -     7.00   96.43   3.9e+03 198440  ! AC009993 Homo sapiens chr
qb_pri0:HUAC002288      +     7.00   95.96   4.1e+03 214633  ! AC002288 Homo sapiens chr
qb_htq12:AC037471       +     7.00   95.86   4.2e+03 217377  ! AC037471 Homo sapiens chr
qb_htq2:AC010736        +     7.00   95.76   4.3e+03 220615  ! AC010736 Homo sapiens chr
qb_htq8:AC022169        +     7.00   94.24   5.2e+03 281383  ! AC022169 Homo sapiens chr

Seq_name: qb_rel:AB019436

Seq_documentation_block:
LOCUS       AB019436      1704 bp    mRNA            ROD       20-JAN-200
DEFINITION  Mus musculus mRNA for Prostate specific transcription factor ets,
            complete cds.
ACCESSION   AB019436
VERSION     AB019436.1  GI:4760536
KEYWORDS    Prostate specific transcription factor ets; mPse.
SOURCE      Mus musculus (strain: C57BL/6) Adult male Tokyo Biochemical cDNA to
            mRNA.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases)
  AUTHORS   Yamada,N., Imai,Y., Miyamoto,H. and Nozaki,M.
  TITLE     Cloning and expression of the mouse Pse gene encoding a novel Ets
            family member
  JOURNAL   Gene 241 (2), 267-274 (2000)
  MEDLINE   20137504
REFERENCE   2  (bases 1 to 1704)
  AUTHORS   Nozaki,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-NOV-1998) to the EMBL/GenBank/DDBJ databases. Masami
            Nozaki, Osaka University, Institute for Microbial Diseases; 3-1,
            Yamadaoka, Suita, Osaka 565-0871, Japan
            (E-mail:mnozaki@biken.osaka-u.ac.jp, Tel:81-6-879-8338,
            Fax:81-6-879-8339)
FEATURES                Location/Qualifiers
     source             1..1704
                        /organism="Mus musculus"
                        /strain="C57BL/6"
                        /db_xref="taxon:10090"
                        /cell_type="Epithelial"
                        /dev_stage="Adult"
                        /sex="male"
                        /tissue_type="Colon"
     gene               362..1339
                        /gene="mPse"
     CDS                362..1339
                        /gene="mPse"
                        /codon_start=1
                        /product="Prostate specific transcription factor ets"
                        /protein_id="BAA77329.1"
                        /db_xref="GI:4760537"
                        /translation="MESAVPCPSNVSTSSLFPVVFEKTSAAYMGLEKQEWS
                        PGPPAIPEQGLSAPYEVPNMYPEGGWVAEVPEAAVTEGHEPEPLCPVPVIC
                        TSSELLLVVDMVVSEVLELTLTATSESLNGTAPGGWCPCNVLFAGISWTPL
                        AGKAPQPLLGKELCAMSEFQFRQAPLGGVVLHAHLLIWKSAAWMKFKTSPG
                        IIHYCAPISEDDWTKGEVDSSCSQQFTHLWQLIKLLLKPHSYGHTFWLNKE
                        VAKLWGVEENEFAMNYLELSKSIELQYYNECLERKTLSQKLVYQFVHPV"
     polyA_site         1704
                        /note="10 a nucleotides"
BASE COUNT       389 a    527 c    491 g    317 t
ORIGIN


Alignment_scores:
                Quality:    8.00              Length:       8
                  Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment Block:
 US-09-126-945B-2_COPY_215_222 x AB019436  ..

 Align seq 1/1  to: AB019436  from: 1  to: 1704
```

```
        1 MetLysGluArgThrSerProGly 8
          ||||||||||||||||||||||||
      974 ATGAAGGAGAGGACCTCGCCTGGG 997

seq name: qb pat1:AX048148

seq documentation block:
LOCUS       AX048148      1894 bp    DNA              PAT       15-DEC-2000
DEFINITION  Sequence 1 from Patent WO0070092.
ACCESSION   AX048148
VERSION     AX048148.1  GI:11876967
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1894)
  AUTHORS   Kaufmann,J. and Xin,H.
  TITLE     Expression of ets domain proteins in cancer
  JOURNAL   Patent: WO 0070092 A 1 23-NOV-2000;
            Chiron Corporation (US)
FEATURES             Location/Qualifiers
     source          1..1894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             416..1423
                     /note="Human JKETS."
                     /codon_start=1
                     /protein_id="CAC19002.1"
                     /db_xref="GI:11876968"
                     /translation="MGSASPGLSSVSPSHLLEPPDLVSPGGLEKAAAQAVSLEKRDWS
                     PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAPCASSREEPPEEPEQCPVIDSQAPA
                     GSLDLVPGGLETLEEHSLEQVSMVVGEVLKDIETACKLLNLTADPMDWSPGNVQKWLL
                     WTEHQYRLPPMGKAFQELAGKELCAMSEEQFRQRSPLGGDVLHAHLDTWKSAAWMKER
                     TSPGATHYCASISEESWTDSEVDSSCSGQPTHEWQFLEELLEPHSYGRFIRWLNKEK
                     GIFKIEDSAQVARLWGIRKNKPAMNYDKLSRSIRQAYYKKGIIRKPTDISQRLVYQFVHP
                     I"
BASE COUNT       368 a    654 c    571 g    302 t
ORIGIN


alignment scores:
            Quality:     8.00              Length:       8
              Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945b-2 copy_215_222 x AX048148   ..

 Align seq 1/1  to: AX048148  from: 1  to: 1894

        1 MetLysGluArgThrSerProGly 8
          ||||||||||||||||||||||||
     1058 ATGAAAGAGCGGACTTCACCTGGG 1081

seq name: qb pr4:AF071538

seq documentation block:
LOCUS       AF071538      1894 bp    mRNA             PRI       10-JAN-2000
DEFINITION  Homo sapiens Ets transcription factor PDEF (PDEF) mRNA, complete
            cds.
ACCESSION   AF071538
VERSION     AF071538.1  GI:4007417
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1894)
  AUTHORS   Oettgen,P., Finger,E., Sun,Z., Akbarali,Y., Thamrongsak,U.,
            Boltax,J., Grall,F., Dube,A., Weiss,A., Brown,L., Quinn,G., Kas,K.,
            Endress,G., Kunsch,C. and Libermann,T.A.
  TITLE     PDEF, a novel prostate epithelium specific ets transcription
            factor, interacts with the androgen receptor and activates
            prostate specific antigen gene expression
  JOURNAL   J. Biol. Chem. 275 (2), 1216-1225 (2000)
  MEDLINE   20092894
REFERENCE   2  (bases 1 to 1894)
  AUTHORS   Libermann,T.A., Oettgen,P., Kunsch,C. and Finger,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUN-1998) Medicine, Beth Israel Deaconess Medical
            Center, 330 Brookline Ave., Boston, MA 02215, USA
FEATURES             Location/Qualifiers
     source          1..1894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6p21.3"
                     /tissue_type="prostate"
     gene            1..1894
                     /gene="PDEF"
     CDS             416..1423
                     /gene="PDEF"
                     /note="prostate epithelium specific"
                     /codon_start=1
                     /product="Ets transcription factor PDEF"
                     /protein_id="AAC95296.1"
                     /db_xref="GI:4007418"
                     /translation="MGSASPGLSSVSPSHLLEPPDLVSPGGLEKAAAQAVSLEKRDWS
                     PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAPCASSREEPPEEPEQCPVIDSQAPA
                     GSLDLVPGGLETLEEHSLEQVSMVVGEVLKDIETACKLLNLTADPMDWSPGNVQKWLL
                     WTEHQYRLPPMGKAFQELAGKELCAMSEEQFRQRSPLGGDVLHAHLDTWKSAAWMKER
                     TSPGATHYCASISEESWTDSEVDSSCSGQPTHEWQFLEELLEPHSYGRFIRWLNKEK
                     GIFKIEDSAQVARLWGIRKNKPAMNYDKLSRSIRQAYYKKGIIRKPTDISQRLVYQFVHP
                     I"
BASE COUNT       368 a    654 c    571 g    302 t
ORIGIN


alignment scores:
            Quality:     8.00              Length:       8
              Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945b-2 copy_215_222 x AF071538   ..

 Align seq 1/2  to: AF071538  from: 1  to: 1894

        1 MetLysGluArgThrSerProGly 8
          ||||||||||||||||||||||||
     1058 ATGAAAGAGCGGACTTCACCTGGG 1081

seq name: qb pr1:AB031549

seq documentation block:
LOCUS       AB031549      1898 bp    mRNA             PRI       20-JAN-2000
DEFINITION  Homo sapiens PSE mRNA for prostate ets, complete cds.
ACCESSION   AB031549
VERSION     AB031549.1  GI:6721497
KEYWORDS    prostate ets.
SOURCE      Homo sapiens adult male prostate epithelium prostate cancer
            cell line:PC-3 cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Yamada,N., Iwari,Y., Miyamoto,H. and Nozaki,M.
  TITLE     Cloning and expression of the mouse Pse gene encoding a novel Ets
            family member
  JOURNAL   Gene 241 (2), 267-274 (2000)
  MEDLINE   20147504
REFERENCE   2  (bases 1 to 1898)
  AUTHORS   Nozaki,M.
  TITLE     Direct Submission
```

```
  JOURNAL    Submitted (25 AUG 2000) to the EMBL/GenBank/DDBJ databases. Misao
             Nozaki, Osaka University, Institute for Microbial Diseases; 3-1,
             Yamadaoka, Suita, Osaka 565-0871, Japan
             (E-mail:mnozaki@biken.osaka-u.ac.jp, Tel:81-6-879-8339,
             Fax:81-6-879-8449)
FEATURES             Location/Qualifiers
     source          1..1898
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="PC-3"
                     /cell_type="prostate cancer"
                     /chromosome="6"
                     /dev_stage="adult"
                     /map="6p21.2-21.3"
                     /sex="male"
                     /tissue_type="prostate epithelium"
                     /note "vector:pOCl8"
     gene            406..1413
                     /gene="PSE"
     CDS             406..1413
                     /gene="PSE"
                     /codon_start=1
                     /product="prostate ets"
                     /protein_id="BAA89543.1"
                     /db_xref="GI:6721498"
                     /translation="MGSASPGLSSVSPSHLLIPPDIVSRGLEKAAAGAVGLERKLWS
                     PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAPGASSKEEPPEEPEQCPVIDSQAPA
                     GSLDLVDGGETLEPHSLEQVQSMVVGPVEPKPLPAVPLLNLGALPDEWGPGNVLEWLL
                     WIEHQYRLPPMGKAFQELACKELCAMSEEQFRQRSPLGGDVLHAHLDWKSAAWMKER
                     TSRGAIHYCASISLSWLDSEVDPSSGFPFWLPLFFLLIEHSSDGFLFWLNKEE
                     GIEKIEDSAQVAPLWGIRKNRPAMNYDKLSRSLRQYYKKGLLRKPLISQRLVYQFVHP
                     I"
BASE COUNT       380 a     651 c     567 g     800 t
ORIGIN


alignment_scores:
            Quality:     8.00              Length:       8
              Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_215_222 x AB041549  ..

 Align seq 1/1  to: AB041549  from: 1  to: 1800

       1 MetLysGluArgThrSerProGly 8
         ||||||||||||||||||||||||
    1048 ATGAAAGAGCGGACTTCACCTGGG 1071

seq_name: gb_pat1:AX048159

seq_documentation_block:
LOCUS        AX048159     3817 bp    DNA              PAT       15-DEC-2000
DEFINITION   Sequence 12 from Patent WO0070092.
ACCESSION    AX048159
VERSION      AX048159.1  GI:11876975
KEYWORDS     .
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 3817)
  AUTHORS    Kaufmann,J. and Xin,H.
  TITLE      Expression of ets domain proteins in cancer
  JOURNAL    Patent: WO 0070092-A 12 23-NOV-2000;
             Chiron Corporation (US)
FEATURES             Location/Qualifiers
     source          1..3817
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note "GSEF-encoding sequence with promoter (Figs. 2A-2B)"
     promoter        1..1380
                     /note "Promoter"
     CDS             1815..2822
                     /note "GSEF coding sequence"
                     /codon_start=1
                     /protein_id="CAC19004.1"
                     /db_xref="GI:11876976"
                     /translation="MGSASPGLSSVSPSHLLIPPDIVSRGLEKAAAGAVGLERKLWS
                     PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAPGASSKEEPPEEPEQCPVIDSQAPA
                     GSLDLVDGGETLEPHSLEQVQSMVVGPVEPKPLPAVPLLNLGALPDEWGPGNVLEWLL
                     WIEHQYRLPPMGKAFQELACKELCAMSEEQFRQRSPLGGDVLHAHLDWKSAAWMKER
                     TSRGAIHYCASISLSWLDSEVDPSSGFPFWLPLFFLLIEHSSDGFLFWLNKEE
                     GIEKIEDSAQVAPLWGIRKNRPAMNYDKLSRSLRQYYKKGLLRKPLISQRLVYQFVHP
                     I"
BASE COUNT       710 a    1026 c     970 g     611 t
ORIGIN


alignment_scores:
            Quality:     8.00              Length:       8
              Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_215_222 x AX048159  ..

 Align seq 1/1  to: AX048159  from: 1  to: 3317

       1 MetLysGluArgThrSerProGly 8
         ||||||||||||||||||||||||
    2457 ATGAAAGAGCGGACTTCACCTGGG 2480

seq_name: gb_pri6:AL157372

seq_documentation_block:
LOCUS        AL157372   150660 bp    DNA              PRI       29-NOV-2000
DEFINITION   Human DNA sequence from clone RP11-375E1 on chromosome 6, complete
             sequence.
ACCESSION    AL157372
VERSION      AL157372.18  GI:11493236
KEYWORDS     HTG.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 150660)
  AUTHORS    Skuce,C.
  TITLE      Direct Submission
  JOURNAL    Submitted (29-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
             CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
             requests: clonerequest@sanger.ac.uk
COMMENT      On Nov 30, 2000 this sequence version replaced gi:11134772.
             During sequence assembly data is compared from overlapping clones.
             Where differences are found these are annotated as variations
             together with a note of the overlapping clone name. Note that the
             variation annotation may not be found in the sequence submission
             corresponding to the overlapping clone, as we submit sequences with
             only a small overlap as described above.
             This sequence has been finished according to sequence map criteria
             as follows. An attempt is made to resolve all sequencing problems,
             such as compressions and repeats, but not necessarily within known
             annotated repeat sequence elements. Where the sequence is
             ambiguous, there is an annotation using the 'unsure' feature key.
             The following abbreviations are used to associate primary accession
             numbers given in the feature table with their source databases:
             Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
             on the WORMPEP database can be found at
             http://www.sanger.ac.uk/Projects/C_elegans/wormpep.shtml. This
             was generated from part of bacterial clone contigs of human
             chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
             Group. Further information can be found at
             http://www.sanger.ac.uk/HGP/Chr6.
             RP11-375E1 is from the library RPCI-11.2 constructed at the Roswell
             Park Cancer Institute by the group of Pieter de Jong. For further
```

```
            details see http://bacpac.med.buffalo.edu/
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RPCI-375E1. It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RPCI-375E1 is at 1 in this sequence. The
            true left end of clone RP3-395O22 is at 150561 in this sequence.
            The true right end of clone RPCI-40,L10 is at 113944 in this
            sequence.
FEATURES             Location/Qualifiers
     source          1..150660
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RPCI-375E1"
                     /clone_lib="RPCI-11.2"
BASE COUNT     45470 a    38887 c    38740 g    27563 t
ORIGIN


alignment scores:
            Quality:     8.00            Length:       8
              Ratio:    1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_215_222 x AL157372/rev   ..

 Align seq 1/1  to reverse of: AL157372  from: 1  to: 150660

       1 MetLysGluArgThrSerProGly 8
         ||||||||||||||||||||||||
  124839 ATGAAAGAGCGGACTTCACCTGGG 124816

seq_name: qb_rol:MUSFRD

seq_documentation_block:
LOCUS        MUSFRD          141 bp    mRNA            ROD       27-JAN-1995
DEFINITION   Mus musculus (clone DE-1) mRNA fragment.
ACCESSION    L49118
VERSION      L49118.1  GI:639780
KEYWORDS     .
SOURCE       Mus musculus female adult hypertrophied heart cDNA to mRNA
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciuroqnathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 141)
  AUTHORS    Al Moustata,A.-E., Dorrance,T. and Chalifour,L.E.
  TITLE      Identification of genes up-regulated in retinoic acid induced
             endothelial cell differentiation
  JOURNAL    Unpublished (1995)
FEATURES             Location/Qualifiers
     source          1..141
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /cell_line="C10"
                     /cell_type="endothelial"
                     /dev_stage="adult"
                     /sex="female"
                     /clone="DE 1"
                     /tissue_type="hypertrophied heart"
                     /note="polyomavirus large T antigen transgenic mouse"
BASE COUNT       52 a      29 c      23 g      37 t
ORIGIN


alignment scores:
            Quality:     7.00            Length:       7
              Ratio:    1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_215_222 x MUSFRD   ..

 Align seq 1/1  to: MUSFRD  from: 1  to: 141

       2 LysGluArgThrSerProGly 8
         ||||||||||||||||||||||
      46 AAAGAGAGAACTTCTCCTGGT 56

seq_name: qb_vi2:HIV3A9LTR

seq_documentation_block:
LOCUS        HIV3A9LTR       688 bp    DNA             VRL       06-JUL-1994
DEFINITION   Human immunodeficiency virus type 1 (clone 3A-9) long terminal
             repeat (LTR).
ACCESSION    L28885
VERSION      L28885.1  GI:455228
KEYWORDS     LTR; long terminal repeat.
SOURCE       Human immunodeficiency virus type 1 blood DNA.
  ORGANISM   Human immunodeficiency virus type 1
             Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
             lentivirus group.
REFERENCE    1  (bases 1 to 688)
  AUTHORS    Michael,N.L., D'Arcy,L.A., Ehrenberg,P.K. and Redfield,R.R.
  TITLE      Naturally occurring genotypes of the human immunodeficiency virus
             type 1 long terminal repeat display a wide range of basal and
             tat induced transcriptional activities
  JOURNAL    J. Virol. 68, 3163-3174 (1994)
  MEDLINE    94202304
FEATURES             Location/Qualifiers
     source          1..688
                     /organism="Human immunodeficiency virus type 1"
                     /proviral
                     /db_xref="taxon:11676"
                     /cell_type="mononuclear cells"
                     /tissue_type="blood"
     LTR             1..688
BASE COUNT      174 a    155 c    212 g    147 t
ORIGIN


alignment scores:
            Quality:     7.00            Length:       7
              Ratio:    1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_215_222 x HIV3A9LTR   ..

 Align seq 1/1  to: HIV3A9LTR  from: 1  to: 688

       1 MetLysGluArgThrSerPro 7
         |||||||||||||||||||
      54 ATGAAAGAGACAACTCACCT 74

seq_name: qb_vi2:HIV3B7LTR

seq_documentation_block:
LOCUS        HIV3B7LTR       693 bp    DNA             VRL       06-JUL-1994
DEFINITION   Human immunodeficiency virus type 1 (clone 3B-7) long terminal
             repeat (LTR).
ACCESSION    L28889
VERSION      L28889.1  GI:455232
KEYWORDS     LTR; long terminal repeat.
SOURCE       Human immunodeficiency virus type 1 blood DNA.
  ORGANISM   Human immunodeficiency virus type 1
             Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
             lentivirus group.
REFERENCE    1  (bases 1 to 693)
  AUTHORS    Michael,N.L., D'Arcy,L.A., Ehrenberg,P.K. and Redfield,R.R.
  TITLE      Naturally occurring genotypes of the human immunodeficiency virus
             type 1 long terminal repeat display a wide range of basal and
             tat induced transcriptional activities
  JOURNAL    J. Virol. 68, 3163-3174 (1994)
  MEDLINE    94202304
```

```
FEATURES             Location/Qualifiers
     source          1. .693
                     /organism="Human immunodeficiency virus type 1"
                     /proviral
                     /db_xref="taxon:11676"
                     /cell type "mononuclear cells"
                     /tissue_type="blood"
     LTR             <1. .>693
BASE COUNT       173 a    155 c    214 g    151 t
ORIGIN


alignment_scores:
            Quality:    7.00            Length:       7
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_215_222 x HIV4B7LTR  ..

 Align Seq 1/1  to: HIV4B7LTR  from: 1  to: 693

       1 MetLysGluArgThrSerPro 7
         ||||||||||||||||||||
     574 ATGAAGGAGAGAACCCCACCT 74

seq_name: gb_ro2:RATCEE1B

seq_documentation_block:
LOCUS        RATCEE1B     1442 bp    mRNA            ROD       29 SEP 1999
DEFINITION  Rattus norvegicus mRNA for carboxyesterase E1, partial cds.
ACCESSION   D00362
VERSION     D00362.1  GI:220705
KEYWORDS    carboxyesterase E1; microsomal carboxyesterase.
SOURCE      Rattus norvegicus (strain:Sprague-Dawley) male liver cDNA to mRNA,
            clone:2-1.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1442)
  AUTHORS   Takami,Y., Morohashi,K., Kawabata,S., Ohno,M. and Omura,T.
  TITLE     Molecular cloning and nucleotide sequence of cDNA of microsomal
            carboxyesterase E1 of rat liver
  JOURNAL   J. Biochem. 104 (5), 801-806 (1988)
  MEDLINE   89174514
COMMENT     Sequence analysis of the cloned cDNAs, 8-1 and 2-1 gave the
            complete amino acid sequence of carboxyesterase E1. There were one
            base difference in the overlap region. So, they sequenced two other
            independent clones which cover this point and found that the
            sequence in question was c in both clones, coinciding with clone
            2-1. They also determined the amino terminal sequence of purified
            carboxyesterase E1 up to the 30th amino acid residue.
FEATURES             Location/Qualifiers
     source          1. .1442
                     /organism="Rattus norvegicus"
                     /strain "Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone "2-1"
                     /sex="male"
                     /tissue_type "liver"
     CDS             190. .1442
                     /codon_start 1
                     /product "carboxyesterase E1"
                     /protein_id "BAA20565.1"
                     /db_xref="GI:4433855"
                     /translation="MWICALVWASLAVCPLWPPDSSPPAVVDTVQGKVLGKYVSLEGF
                     QPVAVFLGVPFAKPPLGSLRFAPPEPAEPWSFVKNTTSYPPMCSQDGVVGKLLADMLS
                     TGKESTLLFSLYCPHCHSDEVLMVFLVELLIKNASWVMVWLNRVNPTFDSCAFLYSQL
                     ENVVVVTIQYRLGIWGFFSTGDEHSRGNWAHLDQLAALRWVQDNIANFGGNPDSVTIF
                     GESAGGVSVSALVLSPLAKNLITRAISESGVALTSVLIRKDMKEVTRQIAVLAGCKTT
                     SSAAMVHCLRQKEPALLETVKKPMFHLAFQPVVDGVVLPDMPEILLEKSFNTVPYIV
                     GFNEDEGTLIPMNLLNFPLSNGQLDQKTAMSVLVASFPEVSELGREAIVFQYTDWLH
                     PEDPAKKNRDALDDMFSDVFFGIPAVLMSRSLRDAGAPTYMYEFQYRPSFVSDMKPKT
     sig_peptide     190. .255
     mat_peptide     256. .1442
                     /product "carboxyesterase E1"
     variation       1214
                     /note "base substitution has occurred at this position in
                     clone 8-1"
                     /replace "t"
BASE COUNT       357 a    366 c    357 g    362 t
ORIGIN


alignment_scores:
            Quality:    7.00            Length:       7
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_215_222 x RATCEE1B  ..

 Align Seq 1/1  to: RATCEE1B  from: 1  to: 1442

       2 LysGluArgThrSerProGly 8
         ||||||||||||||||||||
    1290 AAAGAAGGAGAGAACCCCTGGA 1310

seq_name: gb_ba1:AE001735

seq_documentation_block:
LOCUS        AE001735     9949 bp    DNA             BCT       02 JUN-1999
DEFINITION  Thermotoga maritima section 47 of 136 of the complete genome.
ACCESSION   AE001735 AE060512
VERSION     AE001735.1  GI:4981122
KEYWORDS    .
SOURCE      Thermotoga maritima.
  ORGANISM  Thermotoga maritima
            Bacteria; Thermotogales; Thermotoga.
REFERENCE   1  (bases 1 to 9949)
  AUTHORS   Nelson,K.E., Clayton,R.A., Gill,S.R., Gwinn,M.L., Dodson,R.J.,
            Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,
            McDonald,L., Utterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M.,
            Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A.,
            Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D.,
            Eisen,J.A., Fraser,C.M. et,al.
  TITLE     Evidence for lateral gene transfer between Archaea and bacteria
            from genome sequence of Thermotoga maritima
  JOURNAL   Nature 399 (6734), 323-329 (1999)
  MEDLINE   99287816
REFERENCE   2  (bases 1 to 9949)
  AUTHORS   Nelson,K.E., Clayton,R.A., Gill,S.R., Gwinn,M.L., Dodson,R.J.,
            Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,
            McDonald,L., Utterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M.,
            Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A.,
            Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D.,
            White,O., Salzberg,S.L., Smith,H.O., Venter,J.C. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUN-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .9949
                     /organism "Thermotoga maritima"
                     /db_xref "taxon:2336"
     gene            69. .455
                     /gene "IM0603"
     CDS             69. .455
                     /gene "IM0603"
                     /note "similar to GB:AE000657 percent identity: 99.0%;
                     identified by sequence similarity; putative"
                     /codon_start 1
                     /transl_table 11
                     /product "ribosomal protein S6"
                     /protein_id "AAD35688.1"
                     /db_xref "GI:4981123"
```

```
                     /translation="MAYVEEIYLNGIATNVEIEDIEGNIVEDVRPIIIEVEIKG
                     KVERMGMRKEAYEIEKENEIDYIVIEPKTKQNGIELENEYRVTPELIEWLIPPEIT
                     EKKEPKAGKEKAAAFAIESSEIGSEG"
gene                 455..880
                     /gene="TM0604"
CDS                  455..880
                     /gene="TM0604"
                     /note="similar to GB:AE000657 percent identity: 57.14;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="single-stranded DNA binding protein, putative"
                     /protein_id="AAD35689.1"
                     /db_xref="GI:4981124"
                     /translation="MCTENEIIIPEIVEPIPEYIIIYVIEIIAVEIVIENGI
                     DDAQTTDFFFIVTPTRIAEFAFYIPSEIVIVECLMEMPFWEIPSTEKEVEIEVVAN
                     VVKFMDRKPAETVSETEEELEIPEEDFSSDTFSEDEPDF"
gene                 895..1122
                     /gene="TM0605"
CDS                  895..1122
                     /gene="TM0605"
                     /note="similar to SP:P10806 percent identity: 75.00;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="ribosomal protein S18"
                     /protein_id="AAD35690.1"
                     /db_xref="GI:4981125"
                     /translation="MAYKKEKEFIKKTEKTGTEMEIDSVGKEIIKELSEEIIKNGKEIEK
                     RLTGNCSKHQKMVKVAIKKARQMGLLAYLKI"
gene                 complement(1148..1495)
                     /gene="TM0606"
CDS                  complement(1148..1495)
                     /gene="TM0606"
                     /note="similar to  percent identity: 0.00; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAD35691.1"
                     /db_xref="GI:4981126"
                     /translation="MGMHEIHEHIHEIIKRFIVSILLVGVGNEVSIHEIGSVSIGNLI
                     WVGEEVFLNDARDMIEDLGEVYIFKADDIFGGVMLGGVEVATAGEVFEEWQNSIPDIE
                     EIFFEADTFEDGE"
gene                 1640..2488
                     /gene="TM0607"
CDS                  1640..2488
                     /gene="TM0607"
                     /note="similar to  percent identity: 0.00; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAD35692.1"
                     /db_xref="GI:4981127"
                     /translation="MAYVINIVNGIVVTIDGAAVLVGTIIEIVIFIGELLKIIGY
                     LINTHGHADHIGGNAFLKKMIGSKVIAHKMLKLIENEVELFLTAGAYTTNLIENKI
                     LMAEPSAVDEVIEEGTTTIESVDLEIVPLSGHSTNQIGVAVDGVLFCADAVESEETLR
                     KIPIIVEYNVEEFLEIIIFIEEIGYSYYYVIGTEVIEIGIEVEENKEAVENIENAVL
                     EIIVNPLSTDELMKQLEIKKFFIIENLIYTLYKGIVHAYLSFLLKKLIEKLIKNNV
                     LVWRKG"
gene                 2488..3198
                     /gene="TM0608"
CDS                  2488..3198
                     /gene="TM0608"
                     /note="similar to GB:AE000782 percent identity: 56.60;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAD35693.1"
                     /db_xref="GI:4981128"
                     /translation="MIVNWMFIGFTIVGVIVFTAIFFGFPMYALVAIIVIYMNV
```

```
                     AVIEAYELIFGVGLIYAVFLGIIEVIIGAIVIEVGGYGFGALPIVAVGIFAGVFIVM
                     LIAIGYFGGLIGWAIGMFAIIIGVWFGIIIGWIAIVVSGIFFAPFGPFLFFIGLLI
                     KGLWLMGNIATKLAQLVDNVIFFVGAITGYFSLRTITVFLIWFILFEWLFDYLDIWVV
                     VKGVGWMLNSSEIYLQKE"
gene                 4195..4295
                     /gene="TM0609"
CDS                  4195..4295
                     /gene="TM0609"
                     /note="similar to  percent identity: 0.00; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAD35694.1"
                     /db_xref="GI:4981129"
                     /translation="MSSEKVIVIGYMGGENIKEVIEIVGAIGERGIVIVGWIEEEEL
                     PEKENNIRYIPVKCSNKGNIFQKAVKKKQVDEKIRHLVETIDYDILYMHHFPAIKPLK
                     PFVVAKKRGKKIVYDIHEYHPQNFLSIIPPFLSDLKGFVMWKIEKKQLQLSDLCIEVS
                     EEIKNFVVAKGIPEEKIEIVPNYASMFIEPSPDEEEMEIVLVGIIFENLIDFKKLIK
                     ALVERGFSFRIIGMDSEAFSDVPHIYISFLPYEKMMEELSKSMFSLISYFFIGKEDFN
                     YLYSLPHKFYDSIAAGTPVIVKNSFVSMAKMVEDLGIGVVIDPANVEESLKRIENACQ
                     RYEEILENIRRHQEGFVWDEKKEEEFLERVMG"
gene                 4247..5371
                     /gene="TM0610"
CDS                  4247..5371
                     /gene="TM0610"
                     /note="similar to PID:992974 percent identity: 66.2%;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="lipopolysaccharide biosynthesis protein"
                     /protein_id="AAD35695.1"
                     /db_xref="GI:4981130"
                     /translation="MGKEEGKRISGKSYGLRVLSLVGARPQIIKEAMLHREFKEKGIE
                     EILVHSCQHYDYNMSSVFFEILEIRQPHYNLNVGSGTHGEMTGKIMIEFEKVLLSEKP
                     DLVLVYGDTNTTLSGALVAAKLKIPVAHVEAGLRQHPKDMPEEINKIVTDRVSQLLFC
                     PSKLAVRNLERECITEGVYEVGDVMYDLFLKMEDPEFYDVFEKIGLKENEFVLVTLHR
                     QENVDDPVKLPKIIEQLKRVSKEKKVVFPTHPRTKNRVKEFGLENLLEGMVVIDPVDY
                     LNLMGLVKRCWKVVTDSGGLQKEAYFAGKRAIVVMPDTGWRELVEAGWNKLASEENLF
                     DVIMEEDRSEYPSGLYGDGNAAGKIVEVIEEWFRSLRGIH"
gene                 5407..5577
                     /gene="TM0611"
CDS                  5407..5577
                     /gene="TM0611"
                     /note="similar to  percent identity: 0.00; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAD35696.1"
                     /db_xref="GI:4981131"
                     /translation="MILQIEEGFIEIEEEGIVVEIIEGVGLGLSKIPSVGDEALSMI
                     WSESVRWSSKMW"
gene                 5649..5758
                     /gene="TM0612"
CDS                  5649..5758
                     /gene="TM0612"
                     /note="similar to  percent identity: 0.00; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAD35697.1"
                     /db_xref="GI:4981132"
                     /translation="MIKLEEIMIEIVEVIIEINEEVLAVIIGIIVEIWEVL"
gene                 5816..6187
                     /gene="TM0613"
CDS                  5816..6187
                     /gene="TM0613"
                     /note="similar to GP:Pyro 8 percent identity: 64.71;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
```

```
                     /product "conserved hypothetical protein"
                     /protein_id="AAD35698.1"
                     /db_xref "GI:4981143"
                     /translation="MSAAEIIRRAFRFCINCNRA        AAPAVFAVFAVFLMCFA
                     QAWGYSVPDFLGELSSRFEIPEELMDHALELDKACIPTRYTPDALPSGSPRNRYSRTEA
                     ERLVNYAEKIIRFCEDLLSKI"
     gene            6138..6488
                     /gene="TM0614"
     CDS             6138..6488
                     /gene="TM0614"

alignment_scores:
            Quality:    7.00            Length:       7
              Ratio:   1.000             Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_215_222 x AE001735  ..

 Align Seq 1/1  to: AE001735  from: 1  to: 9949

       2 LysGluArgThrSerProGly 8
         ||||||||||||||||||||
    5388 AAAGAGCGCACAACTCCCGGT 5408

seq_name: gb_in3:CELC23H3

seq_documentation_block:
LOCUS       CELC23H3     39721 bp    DNA             INV       11-APR-2001
DEFINITION  Caenorhabditis elegans cosmid C23H3, complete sequence.
ACCESSION   U80453
VERSION     U80453.1  GI:1707110
KEYWORDS    HTG
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 39721)
  AUTHORS   The C. elegans Genome Sequencing Consortium, Washington University
            Genome Sequencing Center, St. Louis U.S.A. and the Sanger Centre,
            Hinxton, U.K.,C.
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
  JOURNAL   Science 282 (5396), 2012-2019 (1998)
  MEDLINE   99069018
REFERENCE   2  (bases 1 to 39721)
  AUTHORS   Wu,X.
  TITLE     The sequence of C. elegans cosmid C23H3
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 39721)
  AUTHORS   Waterston R
  TITLE     Direct Submission
  JOURNAL   Submitted (29-NOV 1996)
REFERENCE   4  (bases 1 to 39721)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     Submitted by:
                     Genome Sequencing Center
                     Department of Genetics, Washington University
                     St. Louis, MO 63110, USA, and
                     Sanger Centre, Hinxton Hall
                     Cambridge CB10 1RQ, England
                     e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk


            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            WARNING: These data have only had automated annotation
            and have not yet been subjected to manual review of that
            annotation.  We will be manually reviewing this information
            as quickly as possible but at that time this GenBank record
            will be updated and this warning removed.


            NOTES:

            Coding sequences below are predicted from computer analysis, using
            the program GeneFinder(P. Green and L. Hillier, ms in preparation).
FEATURES             Location/Qualifiers
     source          1..39721
                     /organism="Caenorhabditis elegans"
                     /strain="Bristol N2"
                     /db_xref="taxon:6239"
                     /chromosome="II"
                     /clone="CELC23H3"
     gene            2957..5236
                     /gene="C23H3.5"
     CDS             join(2957..3014,3360..4110,4278..4339,4989..5236)
                     /gene="C23H3.5"
                     /note="coded for by C. elegans cDNA yk167a10.3; coded for
                     by C. elegans cDNA yk167a10.5"
                     /codon_start=1
                     /product="Hypothetical protein C23H3.5"
                     /protein_id="AAK31444.1"
                     /db_xref="GI:13592345"
                     /translation="MESNDIRGQLEGALNQFVELEMRLLLSNQVESCFLEELLSNDIII
                     AKIMHPLLKSIVGLEKNPINTEAEAIAQWLIYEQISPHRHLEELFHPEWQMSELVEKYN
                     EMSLQKRNNQLPKFAFNCGGDDVYPLELIKAVRYMDCAYVFFVYLLKPIKESVKPAIG
                     AILLELIHWLNLSLVIVALCSQTHETVVAVLLEMLKVYFIAAGVMMIALLKELLKELN
                     NLLVVPKLHLSPTLTLFKFMFTSNLIAISELLKYNAIIMELAGLNCLRMKISFVLID
                     ILAADLSSPLSTQLQKKKTNNSVQMDMSDLGISISREDMSFLNTLGQGCSSTLELESD
                     PLALLSPNPLPASFSIELENKLKFMNPVAFSTLFLIKLISFLSAIMQLLNSSLDAAGI
                     PKLKKLKEMNKF"
     gene            7032..12798
                     /gene="C23H3.4"
     CDS             join(7032..7255,7411..7511,7588..7988,12189..12451,
                     12711..12798)
                     /gene="C23H3.4"
                     /note="Contains similarity to Pfam domain: PF00222
                     (aminotran_2), Score 136.6, E-value 1.4e-37, N 1; coded
                     for by C. elegans cDNA yk20la1.3; coded for by C. elegans
                     cDNA yk20la1.5; coded for by C. elegans cDNA yk471e08.3;
                     coded for by C. elegans cDNA yk471e08.5"
                     /codon_start=1
                     /protein_id="AAK31446.1"
                     /db_xref="GI:13592347"
                     /translation="MSTIPGCWHPVYIELLVALLAYVVMLNPFFPQQEKLSFELIFRQ
                     KPTTIAFWIFFSVFFFPVFLNPVVAAGRMIEVSTSPGEYLNMASINPLSFFIV
                     KKIFDPAMQLIFKSYGVGSTGPKGFGIVDVHILDLFKELAKFMGGFEAVFSYGFAIVS
                     SAIPAYAKFCIVIFVIPGCVNFAIQFGCLQACPSFVFSYFPDLMPLIFFLFQSQMDKK
                     DKKAASVPFFAVFCLYVNYAGSTLIFFIFFFWSFFVVSVFIFSWSFSVVSSGFG
                     VIFHFNVFMEDVDMVMASLFNALASIGGFCVGRSYVVGHGRLSGLGYCTPSASCIFLLA
                     IAASFAISIIDEEPSKVQKVIFMAINGQKKIQDALSGSKFSIQGCFESPMKHIYYNGE
                     DEEKQLDIFVFITVFTKNHLLLFRARYLDKDELFKIRFSIRVMFQHDLIEEFIQKAVDA
                     IAVVAHEP"
     gene            complement(8468..10677)
                     /gene="C23H3.7"
     CDS             complement(join(8468..8719,8767..10079,10126..10264,
                     10528..10677))
                     /gene="C23H3.7"
                     /note="strong similarity to alpha, alpha, trehalose
                     glucohydrolases"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAK31441.1"
                     /db_xref="GI:13592342"
                     /translation="MREILCAILLIVYVCASLPEFSCENIIDDNADFGIIFHENRVH
                     IELLQEIVSRSISVFQTDKEVAQLIGSDYENGSVLPCDNEIAPGQWMIYCSSKLLQIV
                     MAVQLYPDSELIEVDQPMKENQICKSIMEHEEKFFEVSIESLIKKDVAEEVIDEFIKEG
                     NELDVCDLFDWRPTTEQLANIKDASYQAFAQRLHFIWIQLCRQIKFFVKNDFSKFSLI
```

```
                     YVPYQFILPGGRFREFYYWDAYWTLKGLIASELYSTARMMTLNTAHETEIYGFVPNGG
                     RVYYLKRSQPFFFAPMVYEYYLATGDTQLVADLLFPVIEKEYTFWSERRTVNVEYEHPD
                     LNETLHMFQYKYEFAFLFPPGSFYPQVIPSAFHFLIPKPPEKQEEKSGKAAEKAWQESK
                     WEKNHKDISTIETTNIVPVELNAFLCYNMNIMSLIEYKLTGNTKLRLLWSGSLTNFKLA
                     FTKVFYYVPARKCWYDYNLRKILTHNTIPFASNAVPKESQCTYPLNSQLAVEYVNEMQNS
                     GAFSIPGGTPTESMNELLNQGWDLTNGWGFMNHMITLGLKKSNNFTLQQKAFILAEKWL
                     ETNMQTFNVSIGLMWLKYNVKLTGKKAFSGAYEVVAVGGWCNGAALGLTLVSLGLQS
                     NGPILESLVGSQTISMKSSPDSSTSLPTDLTTTFTTSSSSSTFCGYSNTLTLLTTVFVLY
                     TL"
gene                 13045..15277
                     /gene "C23H3.3"
CDS                  join(13045..13219,13680..13815,13866..13990,14036..14118,
                     14547..14648,14703..14836,14896..15277)
                     /gene "C23H3.3"
                     /note "Contains similarity to Pfam domain: PF00013
                     (KH-domain), Score 12.8, E value 0.089, N 1"
                     /codon_start 1
                     /evidence not_experimental
                     /protein_id "AAK31443.1"
                     /db_xref "GI:13592344"
                     /translation "MAATPPLSAGTYSTNGKKYRKTAYNGKAQTGFGDLNGTGTTADTP
                     DTYEDEASADCFAPPGLPELKKRAAKAKTVAEAVETSSKEPEPTEDDPKTKVTGGKEWT
                     SSTNVAPCFTGKLTGTNKFTLNSLENETQCTVKTTPPNPNTACEESSTVSLDTVQPTL
                     DRLETFTDDSKKTAKVNHFVALFCDQHEVQENENTFFKQMVMESDHFDSSCKNSQLFTK
                     PTRLHLTLSVARTFDDMDLQKAVGAFETLEKETPQTKDSKPLIADTQGTDMMNDDPSQ
                     VFVLYAKVKGDEVQEVANYVNRELTRSCVSSKNEHDNGSDAVKLHMTLMNSKYVTQSE
                     KSGKSKEAAITFAKLVLEDLFLSVPTTTLFEVNEFVKETEQEL"
gene                 16688..19515
                     /gene "C23H3.2"
CDS                  join(16688..16758,16901..16962,17006..17058,18451..18535,
                     18580..18776,18824..18946,18994..19385,19449..19515)
                     /gene "C23H3.2"
                     /note "coded for by C. elegans cDNA yk701h11.5; coded for
                     by C. elegans cDNA yk701h11.3"
                     /codon_start 1
                     /product "Hypothetical protein C23H3.2"
                     /protein_id "AAK31447.1"
                     /db_xref "GI:13592348"
                     /translation "MINSIVLTGSPKSNKSSQAVTLAKSLSSEKGSESPKSLPKYHPC
                     GSGSLPGGATVLTNNKPDMKKYKQKLDNKTGIVLLMCLLIMFVLTGIYGIMHSSVHD
                     ELTTYVNRTHGTASELKNAMDENYPEENQTNVDKRCRVWMTVEELEDYAYLIPKTNSF
                     VWNLFFWNLLLSIFLFPIVLATHMGWIAGNGAKFAYRIVLLIGLLFCIAQFLYLVHPI
                     FWGAVREPGMLDRLFLEAYPRDQYGTDSTGKREACTEFKPHELVQLDLQQQKCTFKMKN
                     SLLPTYSFLLLIFIDLFPFLFGAFIYAWSAWIKNSQVVVRIARQRVELNNSKKVPFPTQ
                     NIYSPPERNVKLVEV"
gene                 25054..26429
                     /gene "C23H3.1"
CDS                  join(25054..25248,25299..25553,25829..26015,26093..26429)
                     /gene "C23H3.1"
                     /codon_start 1
                     /evidence not_experimental
                     /product "Hypothetical protein C23H3.1"
                     /protein_id "AAK31442.1"
                     /db_xref "GI:13592343"
                     /translation "MLFCIQPSSSSSGGGKADENVKTLMAKVENLEGDISLESRSG
                     VTALGQHYVLVTFFKMFGGYFTHMTIANMKNPFFISIPPKVSSGAQSIPPSPPDLTSPL
                     TTVLGRPGMPSFDQPPIGSQFPVCKELIQLLEVPVGVHDRQDKYLEKGDEVPCEVNVS
                     GVKFYHSGIYAGDGMCYHFVCDAQESESFADALAVFSGASAHVVYDIWFEFVYALVEV
                     SDVPPKIFRASHPLICPSCEQVVFYAFHLQKELENYDIRPCNCQHFSSPQSFCVPFSY
                     DMTSNFKYLACTVLKPTSTVVNAMTPPNRDRSSFASSSISS"
gene                 complement(27112..28800)
                     /gene "C23H3.8"
CDS                  complement(join(27112..27259,27319..27372,27481..27611,
                     27666..27985,28416..28800))
                     /gene "C23H3.8"
                     /note "coded for by C. elegans cDNA yk437q11.3; coded for
                     by C. elegans cDNA yk497e11.5; coded for by C. elegans cDNA
                     CEESQ58F"
                     /codon_start 1
                     /product "Hypothetical protein C23H3.8"
                     /protein_id "AAK31445.1"
                     /db_xref "GI:13592346"
                     /translation "MKLEISELEMEPMSPWCSPVKTPTVGLFIGIFDFLFLTATFLKS
                     VGLLQLYGFNETATAAVSSIGVLYAVHLLCTLAFWYGLAKQKSSWLVFKVLKGTIVL
                     ICVATLTLAFPVSSNSFIGNATAIELAKESVFLDQTDSVVITRASPGHMGNMNFFA
                     GKFIYAYSMEYIGHVIFILLFLPYPKLSCNVEVWFIFGHTQHDLLSVLVHNVAGYGI
                     MATVGGLIFGKLILSLKKRVMTELSVWENATKKQNANLIVCMALWIVFGHGQFTEVLP
                     NYEGFAVCTFLLAQGHLLNQALLSFLGLPHLCLQNFLPFFLVCTMY
                     D*SRFKQLFNS"
     gene            complement(29554..32668)
                     /gene "C23H3.9"
     CDS             complement(join(29554..29614,29664..29768,29814..30255,

alignment scores:
            quality:     7.00              Length:       7
              Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_215_222 x CELC23H3   ..

 Align seq 1/1  to: CELC23H3  from: 1  to: 39721

        1 MetLysGluArgThrSerPro 7
          |||||||||||||||||||||
    27211 ATGAAAGAAAGAACATCTCCA 27235

seq_name: qb_pr2:AC005756

Accession_block:
LOCUS       AC005756     43299 bp    DNA              PRI       02-OCT-1998
DEFINITION  Homo sapiens chromosome 19, fosmid 89347, complete sequence.
ACCESSION   AC005756
VERSION     AC005756.1  GI:3688091
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 43299)
  AUTHORS   Lamerdin,J.E., McCready,P.M., Skewes-Cox,P., Viswanathan,V.,
            Berkhart-Schultz,K., Gordon,L., Dias,J., Ramirez,M., Stilwagen,S.,
            Pham,H., Velasco,N., Do,L., Regala,W., Terry,A., Gaines,J.,
            Danaatan,L., Poundstone,P., Christensen,M., Georgescu,A., Avila,J.,
            Lin,S., Allix,C., Andreiso,T., Trankheim,M., Amiss Keller,G.,
            Coetzeis,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G.,
            Krummiller,B., Arellano,A., Saunders,C., Ow,D., Nolan,M., Trout,S.,
            Kobayashi,A., Olsen,A.S. and Carrano,A.V.
  TITLE     Sequence analysis of a 3.5 Mb contig in human 19p13.3 containing a
            serine protease gene cluster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 43299)
  AUTHORS   Lamerdin,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-OCT-1998) Joint Genome Institute, Lawrence Livermore
            National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
COMMENT     Map and sequence oriented from p telomere to centromere. Fosmid
            89347 is separated from cosmid F23259 to the left by approximately
            1 Kb, and is separated from cosmid R32202 to the right by a gap of
            undetermined size. Additional chromosome 19 map and sequence
            information may be obtained at:
            http://www.bio.llnl.gov/bbrp/genome/genome.html.
FEATURES             Location/Qualifiers
     source          1..43299
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="fos89347"
                     /chromosome="19"
                     /map="19p13.3 between CD034 and D19S842"
                     /cell_line="5HL2-B"
                     /clone_lib="LLNL-C04 W chromosome 19 specific fosmid
                     library"
                     /note="Fosmid library constructed at LLNL from flow sorted
                     chromosomes from hybrid 5HL2-B, which carries chromosome
                     19 as its only human chromosome."
     repeat_region   144..184
```

```
                     /rpt_family="MIR"
     repeat_region   complement(318..384)
                     /rpt_family "(TA)n"
     repeat_region   complement(385..686)
                     /rpt_family "AluY"
     repeat_region   complement(696..971)
                     /rpt_family "AluJo"
     repeat_region   1265..1542
                     /rpt_family "AluSx"
     CDS             join(1548..1768,2085..2226)
                     /note="Hypothetical human protein most similar to
                     elongation factor SIII p15 subunit [Rattus norvegicus];
                     Hypothetical protein shares 62% residue identity with
                     L29259) elongation factor SIII p15 subunit [Rattus
                     norvegicus] over length of protein, and 59% residue
                     identity with (AB007693) Elongin C [Drosophila
                     melanogaster] over length of protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /product "tos39347_1"
                     /protein_id "AAC22620.1"
                     /db_xref "GI:3688092"
                     /translation="MNGIEESCINSGLAMLGLVSPTLGLVEEHLSKLVEQH
                     VLSGPGQSAFNFTLQVNFWFIPPQCVEKQAKAQIYFTYKVHYSNSSIFIPFLIALFV
                     LMASDFLDIYIFLFFLFF"
     repeat_region   1787..2087
                     /rpt_family="AluJo"
     repeat_region   complement(2205..2508)
                     /rpt_family "AluY"
     repeat_region   complement(2747..3042)
                     /rpt_family "AluSq"
     repeat_region   complement(3282..3402)
                     /rpt_family "(GGAA)n"
     repeat_region   complement(3404..3538)
                     /rpt_family "(GA)n"
     repeat_region   complement(3541..3585)
                     /rpt_family="(GAAA)n"
     repeat_region   complement(3587..3888)
                     /rpt_family="AluJo"
     repeat_region   4627..4760
                     /rpt_family="AluJo"
     repeat_region   4808..4886
                     /rpt_family "AluJ/FRAM"
     repeat_region   4891..5194
                     /rpt_family="AluSq"
     repeat_region   5198..5470
                     /rpt_family "AluJo"
     repeat_region   complement(5551..5853)
                     /rpt_family="AluSx"
     repeat_region   5890..6135
                     /rpt_family "MLT1E"
     repeat_region   complement(6349..6656)
                     /rpt_family "AluSx"
     repeat_region   complement(6660..6962)
                     /rpt_family "AluJb"
     misc_feature    7184..7253
                     /note="predicted exon, program: grail2exons_human_1.x
                     frame: 1, quality: excellent, score: 84.500"
     repeat_region   complement(7874..8128)
                     /rpt_family="AluSq"
     repeat_region   8141..8344
                     /rpt_family="MER20"
     repeat_region   8573..8765
                     /rpt_family="MER20"
     repeat_region   8804..9107
                     /rpt_family="AluSq"
     repeat_region   9110..9412
                     /rpt_family "AluSx"
     misc_feature    complement(9654..9780)
                     /note="predicted exon, program: grail2exons_human_1.x,
                     frame: 0, quality: good, score: 60.000"
     repeat_region   complement(10687..11107)
                     /rpt_family "MLT1D"

     repeat_region   11533..11938
                     /rpt_family "AluSx"
     repeat_region   11946..12245
                     /rpt_family "AluSx"
     repeat_region   complement(12742..13020)
                     /rpt_family "AluSq1"
     repeat_region   complement(13667..13965)
                     /rpt_family "AluSx"
     repeat_region   14099..14254
                     /rpt_family "MIR"
     misc_feature    14658..14796
                     /note "predicted exon, program: grail2exons_human_1.x,
                     frame: 0, quality: good, score: 62.000"
     repeat_region   14932..15135
                     /rpt_family "(TA)n"
     repeat_region   complement(15136..15421)
                     /rpt_family "AluSx"
     repeat_region   15428..15593
                     /rpt_family "MER1A"
     repeat_region   15597..15758
                     /rpt_family "AluSx"
     repeat_region   15760..15819
                     /rpt_family "MER1A"
     repeat_region   complement(15821..16122)
                     /rpt_family "AluSx"
     repeat_region   16125..16371
                     /rpt_family "MER1A"
     repeat_region   16372..16661
                     /rpt_family "L1"
     misc_feature    16598..16707
                     /note "predicted exon, program: grail2exons_human_1.x,
                     frame: 0, quality: good, score: 69.000"
     repeat_region   16822..16948
                     /rpt_family "(TA)n"
     repeat_region   16947..17068
                     /rpt_family "(CATA)n"
     repeat_region   17141..17266
                     /rpt_family "(CATA)n"
     repeat_region   17358..17466
                     /rpt_family "(TA)n"
     repeat_region   17471..17560
                     /rpt_family "AluJo/FRAM"
     repeat_region   complement(17586..17619)
                     /rpt_family "AT_rich"
     repeat_region   complement(17620..17939)
                     /rpt_family "AluJo"
     repeat_region   complement(17940..17997)
                     /rpt_family="(TA)n"
     repeat_region   complement(17998..18291)
                     /rpt_family "AluSx"
     misc_feature    complement(18587..18626)
                     /note "predicted exon, program: grail2exons_human_1.x,
                     frame: 1, quality: excellent, score: 95.000"
     repeat_region   complement(20188..20395)
                     /rpt_family "MER20"
     repeat_region   20451..20594
                     /rpt_family "(CA)n"
     repeat_region   20598..20647
                     /rpt_family "(CA)n"
     repeat_region   20640..20806
                     /rpt_family "(CA)n"
     repeat_region   complement(21281..21578)
                     /rpt_family "AluY"
     repeat_region   complement(21679..21970)
                     /rpt_family "AluJo"
     repeat_region   complement(22138..22235)
                     /rpt_family "(GAAA)n"
     repeat_region   complement(22237..22537)
                     /rpt_family "AluSx"
     repeat_region   22566..22613
                     /rpt_family "(TA)n"
     repeat_region   22627..22787
                     /rpt_family "MER82"
```

```
     repeat_region   complement(22788, 22954)
                     /rpt_family "AluJo/FRAM"
     repeat_region   complement(22957, 23248)
                     /rpt_family "AluSp"
     misc_feature    complement(23239, 23361)
                     /note "predicted exon, program: grail2exons_human 1.3,
                     frame: 0, quality: good, score: 74.000"
     repeat_region   23362, 23668

alignment scores:
              Quality:     7.00              Length:       7
                Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2_COPY_215_222 x AC005756  ..

 Align seq 1/1  to: AC005756  from: 1  to: 43299

       2 LysGluArgThrSerProGly 8
         ||||||||||||||||||||||
   19182 AAGGAAAGAACGTCTCCAGGC 19202

Seq name: qb_htg17:AC080145

seq documentation block:
LOCUS       AC080145    110697 bp    DNA              HTG       14-FEB-2001
DEFINITION  Homo sapiens chromosome 15 clone CTD 2298D17 map 15, WORKING DRAFT
            SEQUENCE, 24 unordered pieces.
ACCESSION   AC080145
VERSION     AC080145.2  GI:12830209
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 110697)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone CTD 2298D17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 110697)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,B., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Beda,F., Bequiavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
            FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Glude,S., Goyette,M.,
            Graham,L., Grand Pierre,N., Hagos,B., Heaford,A., Horton,L.,
            Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaBocque,K.,
            Lamazares,K., Landers,T., Lebkowsky,J., Levine,M., Liao,C., Liu,S.,
            Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Menezes,L., Mihova,T., Mlenga,V.,
            Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
            O'Donnell,P., O'Neil,D., Oliver,T.M., Oliver,J., Peterson,E.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Richcreek,M., Riley,R.,
            Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
            Sonqnez,C., Spencer,B., Stange Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
            Tirrell,A., Travers,M., Triqilio,J., Vassiliev,R., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,C., Zainoun,J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26 SEP 2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 14, 2001 this sequence version replaced gi:10305247.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
                  ----- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www.seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
                        Project Information
              Center project name: LI0681
              Center clone name: 2298 D 17
                        Summary Statistics
              Sequencing vector: Plasmid; n/a; 100% of reads
              Chemistry: Dye terminator Big Dye; 100% of reads
              Assembly program: Phrap; version 0.960731
              Consensus quality: 99944 bases at least Q40
              Consensus quality: 105546 bases at least Q30
              Consensus quality: 107655 bases at least Q20
              Insert size: 115000; agarose fp
              Insert size: 108897; sum of contigs
              Quality coverage: 4.6 in Q20 bases; agarose fp
              Quality coverage: 3.8 in Q20 bases; sum of contigs

            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 24 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     739: contig of 739 bp in length
            *     740  839: gap of     100 bp
            *     840 1982: contig of 1143 bp in length
            *    1983 2082: gap of     100 bp
            *    2083 4194: contig of 2112 bp in length
            *    4195 4294: gap of     100 bp
            *    4295 4180: contig of 886 bp in length
            *    4181 4280: gap of     100 bp
            *    4281 5594: contig of 1314 bp in length
            *    5595 5694: gap of     100 bp
            *    5695 7447: contig of 1753 bp in length
            *    7448 7547: gap of     100 bp
            *    7548 8770: contig of 1223 bp in length
            *    8771 8870: gap of     100 bp
            *    8871 10132: contig of 1262 bp in length
            *   10133 10232: gap of     100 bp
            *   10233 11410: contig of 1078 bp in length
            *   11411 11410: gap of     100 bp
            *   11411 13682: contig of 2272 bp in length
            *   13683 13782: gap of     100 bp
            *   13783 16290: contig of 2508 bp in length
            *   16291 16390: gap of     100 bp
            *   16391 17939: contig of 1549 bp in length
            *   17940 18039: gap of     100 bp
            *   18040 20281: contig of 2242 bp in length
            *   20282 20381: gap of     100 bp
            *   20382 22812: contig of 2431 bp in length
            *   22813 22912: gap of     100 bp
            *   22913 24536: contig of 1624 bp in length
            *   24537 24636: gap of     100 bp
            *   24637 28872: contig of 4241 bp in length
            *   28873 28972: gap of     100 bp
            *   28973 36783: contig of 7811 bp in length
            *   36784 36883: gap of     100 bp
            *   36884 41605: contig of 4717 bp in length
            *   41606 41705: gap of     100 bp
            *   41706 47287: contig of 5582 bp in length
            *   47288 47387: gap of     100 bp
            *   47388 54803: contig of 7416 bp in length
            *   54804 54903: gap of     100 bp
            *   54904 63579: contig of 8676 bp in length
            *   63580 63679: gap of     100 bp
            *   63680 72766: contig of 9087 bp in length
            *   72767 72866: gap of     100 bp
            *   72867 85422: contig of 12556 bp in length
            *   85423 85522: gap of     100 bp
            *   85523 110697: contig of 25175 bp in length.
FEATURES             Location/Qualifiers
     source          1..110697
                     /organism "Homo sapiens"
```

```
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map "15"
                     /clone="CTD-2298D17"
                     /clone_lib "CITB Human BAC"
     misc_feature    1. .739
                     /note="assembly_fragment"
     misc_feature    840. .1982
                     /note="assembly_fragment"
     misc_feature    2083. .4194
                     /note="assembly_fragment"
     misc_feature    3295. .4180
                     /note="assembly_fragment"
     misc_feature    4281. .5594
                     /note="assembly_fragment"
     misc_feature    5695. .7447
                     /note="assembly_fragment"
     misc_feature    7548. .8770
                     /note="assembly_fragment"
     misc_feature    8871. .10132
                     /note="assembly_fragment"
     misc_feature    10233. .11310
                     /note="assembly_fragment"
     misc_feature    11411. .13682
                     /note="assembly_fragment"
     misc_feature    13783. .16290
                     /note="assembly_fragment"
     misc_feature    16391. .17939
                     /note="assembly_fragment"
     misc_feature    18040. .20281
                     /note="assembly_fragment"
     misc_feature    20382. .22812
                     /note="assembly_fragment"
     misc_feature    22913. .24536
                     /note="assembly_fragment"
     misc_feature    24637. .28877
                     /note="assembly_fragment"
     misc_feature    28978. .36788
                     /note="assembly_fragment"
     misc_feature    36889. .41605
                     /note="assembly_fragment"
     misc_feature    41706. .47287
                     /note="assembly_fragment"
     misc_feature    47388. .54803
                     /note="assembly_fragment"
     misc_feature    54904. .63579
                     /note="assembly_fragment"
     misc_feature    63680. .72766
                     /note="assembly_fragment"
     misc_feature    72867. .85422
                     /note="assembly_fragment"
     misc_feature    85523. .110697
                     /note="assembly_fragment"
BASE COUNT    33553 a   20748 c   21247 g   32952 t    2202 others
ORIGIN


Alignment_scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment_block:
 US-09-126-945B-2_COPY_215_222 x AC080145   ..

 Align seq 1/1  to: AC080145  from: 1  to: 110697

       1 MetLysGluArgThrSerPro 7
         ||| ||||||||||||||||| 
   77689 ATGAAAGAGAGAACCTCTCCA 77709

seq_name: qb_htq19:ALI39159
```

```
Sequence alignment block:
LOCUS       ALI39159    128490 bp    DNA             HTG       28-JAN-2001
DEFINITION  Homo sapiens chromosome 1 clone RP5-894H24, *** SEQUENCING IN
            PROGRESS ***, 14 unordered pieces.
ACCESSION   ALI39159
VERSION     ALI39159.5  GI:9796347
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 128490)
  AUTHORS   Humb,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 11, 2000 this sequence version replaced gi:4230649.
             ----------  Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@Sanger.ac.uk
              ------------ Project Information
            Center project name: dJ894H24
              ----- ------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; U08752; 100% of reads
            Chemistry: Dyeterminator Big Dye; 100% of reads
            Consensus quality: 117159 bases at least Q40
            Consensus quality: 121601 bases at least Q30
            Consensus quality: 124636 bases at least Q20
            Insert size: 127190, sum of contigs
            Insert size: 146739; 0.7% error; agarose-fp
            Quality coverage: 3.85x in Q20 bases, sum of contigs quality
            coverage: 3.1%x in Q20 bases; agarose-fp
            ------ ----- .
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     6945: contig of 6945 bp in length
            *     6946  7045: gap of      100 bp
            *     7046    20449: contig of 13404 bp in length
            *    20450  20549: gap of      100 bp
            *    20550    33676: contig of 13127 bp in length
            *    33677  33776: gap of      100 bp
            *    33777    43983: contig of 10207 bp in length
            *    43984  44083: gap of      100 bp
            *    44084    49814: contig of 5731 bp in length
            *    49815  49914: gap of      100 bp
            *    49915    55620: contig of 5706 bp in length
            *    55621  55720: gap of      100 bp
            *    55721    77709: contig of 21989 bp in length
            *    77710  77809: gap of      100 bp
            *    77810    81041: contig of 3232 bp in length
            *    81042  81141: gap of      100 bp
            *    81142    83691: contig of 2550 bp in length
            *    83692  83791: gap of      100 bp
            *    83792    94399: contig of 10608 bp in length
            *    94400  94499: gap of      100 bp
            *    94500   110766: contig of 16267 bp in length
            *   110767 110866: gap of      100 bp
            *   110867   114954: contig of 4088 bp in length
            *   114955 115054: gap of      100 bp
            *   115055   121888: contig of 6834 bp in length
            *   121889 121988: gap of      100 bp
            *   121989   128490: contig of 6502 bp in length.
FEATURES             Location/Qualifiers
```

```
     source          1..128490
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /chromosome "1"
                     /clone "RP5-894H24"
                     /clone_lib "RPCI-5"
     misc_feature    1..6945
                     /note "assembly_fragment:00846
                     fragment_chain:1"
     misc_feature    7046..20449
                     /note "assembly_fragment:00948
                     fragment_chain:1"
     misc_feature    20550..34676
                     /note "assembly_fragment:00070
                     fragment_chain:1"
     misc_feature    34777..43983
                     /note "assembly_fragment:00893
                     fragment_chain:1"
     misc_feature    44084..49814
                     /note "assembly_fragment:00944
                     fragment_chain:2"
     misc_feature    49915..55620
                     /note "assembly_fragment:00041
                     fragment_chain:2"
     misc_feature    55721..77709
                     /note "assembly_fragment:00688
                     fragment_chain:2"
     misc_feature    77810..81041
                     /note "assembly_fragment:00584
                     fragment_chain:2"
     misc_feature    81142..83691
                     /note "assembly_fragment:00076
                     fragment_chain:3"
     misc_feature    83792..94399
                     /note "assembly_fragment:00799
                     fragment_chain:3"
     misc_feature    94500..110766
                     /note "assembly_fragment:00387
                     fragment_chain:4"
     misc_feature    110867..114954
                     /note "assembly_fragment:01071
                     fragment_chain:4"
     misc_feature    115055..121888
                     /note "assembly_fragment:00439
                     fragment_chain:5"
     misc_feature    121989..128490
                     /note "assembly_fragment:00318
                     fragment_chain:5"
BASE COUNT     31910 a   31655 c   31963 g   31655 t    130 others
ORIGIN


Alignment scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-126-945B-2-COPY_215_222 x AL139159/rev  ..

 Align seq 1/1  to reverse of: AL139159  from: 1  to: 128490

        2 LysGluArgThrSerProGly 8
          ||||||||||||||||||||||
   59099 AAGGAACGTACATCCTGGC 59079
```

[illegible]

Date: Sep 14, 2001, 6:48 PM

About: Results were produced by the Smith-Waterman [illegible]
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL frame+_p2n.model -DEV xlp
-Q [illegible] -DB=N_Geneseq_0601 -QFMT=fastap -SUFFIX olip2n.rmt -GAPOP 4.500
-GAPEXT 0.050 -MINMATCH=0.100 -LOOPCI 0.000 -LOOPEXT=0.000
-QGAPOP 4.500 -QGAPEXT 0.050 -XGAPOP=60.000 XGAPEXT 60.000
-FGAPOP 6.000 FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
-DELOP=6.000 -DELEXT 7.000 -START 1 -MATRIX clip
-TRANS human40.cdi -LIST=45 -DOCALIGN 200 -THR_SCORE quality
-THR_MIN 1 -ALIGN 15 -MODE LOCAL -OUTFMT=pir NORM ext MINLEN 8
-MAXLEN 2000000000 -USER=US09126945_2COPY_215_222 NCPU=6 -ICPU 6
-LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-126-945B-2_COPY_215_222
Query length: 8
Database: N_Geneseq_0601:*
Database sequences: 740101
Database length: 413950809
Search time (sec): 680.150000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Identities			
/SIDSI/seqdata/geneseq/geneseqp/NA2000.DAT:AAA06617	+		8.00	153.96	1.42		278	1
/SIDSI/seqdata/geneseq/geneseqp/NA2000.DAT:AAF21828	+		8.00	141.73	4.42		2082	1
/SIDSI/seqdata/geneseq/geneseqp/NA2000.DAT:AAC85641	+		8.00	135.59	6.66		1041	1
/SIDSI/seqdata/geneseq/geneseqp/NA2001.DAT:AAC83261	+		8.00	140.98	6.99		1804	1
/SIDSI/seqdata/geneseq/geneseqp/NA2001.DAT:AAC83266	+		8.00	147.14	11.38		3217	1
/SIDSI/seqdata/geneseq/geneseqp/NA1999.DAT:AAY89435	+		6.00	118.14	130.88		248	1
/SIDSI/seqdata/geneseq/geneseqp/NA1993.DAT:AAY42335	+		6.00	127.42	141.63		276	1
/SIDSI/seqdata/geneseq/geneseqp/NA2000.DAT:AAY78838	+		6.00	126.92	152.71		297	1
/SIDSI/seqdata/geneseq/geneseqp/NA1999.DAT:AAZ13159	+		6.00	116.85	154.38		300	1
/SIDSI/seqdata/geneseq/geneseqp/NA2000.DAT:AAA00274	+		6.00	116.85	154.38		300	1
/SIDSI/seqdata/geneseq/geneseqp/NA2000.DAT:AAA00351	+		6.00	116.85	154.38		300	1
/SIDSI/seqdata/geneseq/geneseqp/NA2000.DAT:AAA00344	+		6.00	116.85	154.38		300	1
/SIDSI/seqdata/geneseq/geneseqp/NA1992.DAT:AAL20365	+		6.00	116.54	156.73		312	1
/SIDSI/seqdata/geneseq/geneseqp/NA2000.DAT:AAC86774	+		6.00	115.81	176.48		350	1
/SIDSI/seqdata/geneseq/geneseqp/NA2000.DAT:AAY70149	+		6.00	115.81	176.48		350	1
/SIDSI/seqdata/geneseq/geneseqp/NA1996.DAT:AAY94416	+		6.00	114.95	215.62		427	1
/SIDSI/seqdata/geneseq/geneseqp/NA2000.DAT:AAC91236	+		6.00	112.87	257.11		510	1
/SIDSI/seqdata/geneseq/geneseqp/NA2000.DAT:AAA02946	+		6.00	111.41	290.03		575	1
/SIDSI/seqdata/geneseq/geneseqp/NA1998.DAT:AAV17541	+		6.00	111.34	292.81		577	1
/SIDSI/seqdata/geneseq/geneseqp/NA1999.DAT:AAY15863	+		6.00	111.07	324.04		705	1
/SIDSI/seqdata/geneseq/geneseqp/NA2002.DAT:AAB98200	+		6.00	111.01	326.42		711	1
/SIDSI/seqdata/geneseq/geneseqp/NA1998.DAT:AAV52470	+		6.00	110.11	366.70		818	1
/SIDSI/seqdata/geneseq/geneseqp/NA2000.DAT:AAZ46196	+		6.00	110.06	369.05		819	1
/SIDSI/seqdata/geneseq/geneseqp/NA2000.DAT:AAA03514	+		6.00	109.86	378.42		844	1
/SIDSI/seqdata/geneseq/geneseqp/NA2000.DAT:AAA06161	+		6.00	109.54	390.22		1112	1
/SIDSI/seqdata/geneseq/geneseqp/NA2000.DAT:AAY14274	+		6.00	107.62	501.96		1565	1
/SIDSI/seqdata/geneseq/geneseqp/NA2000.DAT:AAC62744	+		6.00	107.07	541.50		1974	1
/SIDSI/seqdata/geneseq/geneseqp/NA2000.DAT:AAZ73667	+		6.00	106.96	555.20		2138	1
/SIDSI/seqdata/geneseq/geneseqp/NA2000.DAT:AAC62754	+		6.00	106.58	577.76		2298	1
/SIDSI/seqdata/geneseq/geneseqp/NA2000.DAT:AAC40904	+		6.00	105.53	656.26		1590	1
/SIDSI/seqdata/geneseq/geneseqp/NA1996.DAT:AAX85746	+		6.00	105.30	678.80		1658	1
/SIDSI/seqdata/geneseq/geneseqp/NA2000.DAT:AAV17467	+		6.00	105.25	682.58		1733	1
/SIDSI/seqdata/geneseq/geneseqp/NA2000.DAT:AAY54531	+		6.00	104.92	713.23		1750	1
/SIDSI/seqdata/geneseq/geneseqp/NA1998.DAT:AAV33216	+		6.00	104.92	713.23		1750	1
/SIDSI/seqdata/geneseq/geneseqp/NA2000.DAT:AAY54535	+		6.00	104.92	713.23		1750	1
/SIDSI/seqdata/geneseq/geneseqp/NA2000.DAT:AAC91174	+		6.00	104.77	726.65		1788	1
/SIDSI/seqdata/geneseq/geneseqp/NA1997.DAT:AAT72191	+		6.00	104.33	760.74		1908	1
/SIDSI/seqdata/geneseq/geneseqp/NA1997.DAT:AAT62287	+		6.00	104.32	760.91		1911	1
/SIDSI/seqdata/geneseq/geneseqp/NA2000.DAT:AAZ25576	+		6.00	104.31	770.80		1914	1
/SIDSI/seqdata/geneseq/geneseqp/NA1996.DAT:AAX85747	+		6.00	104.12	827.86		2060	1
/SIDSI/seqdata/geneseq/geneseqp/NA2000.DAT:AAV16524	+		6.00	101.91	827.11		2221	1
/SIDSI/seqdata/geneseq/geneseqp/NA1998.DAT:AAV29525	+		6.00	101.83	1.1e+03		2764	1
/SIDSI/seqdata/geneseq/geneseqp/NA1998.DAT:AAV17358	+		6.00	101.83	1.1e+03		2764	1
/SIDSI/seqdata/geneseq/geneseqp/NA1996.DAT:AAI19028	+		6.00	101.82	1.1e+03		2765	1

seq_name: /SIDSI/seqdata/geneseq/geneseqp/NA2000.DAT:AAA06617

seq_documentation_block:
ID AAA06617 standard; cDNA; 278 BP.
XX
AC AAA06617;
XX
DT 13-JUN-2000 (first entry)
XX
DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:99.
XX
KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW immunogenic; cytostatic; vaccine; ss.
XX
OS Homo sapiens
XX
PN WO200004149-A2.
XX
PD 27-JAN-2000.
XX
PF 14-JUL-1999; 99WO-US15838.
XX
PR 14-JUL-1998; 98US-0115454.
PR 14-JUL-1998; 98US-0116134.
PR 23-SEP-1998; 98US-0159812.
PR 23-SEP-1998; 98US-0159822.
PR 15-JAN-1999; 99US-0232149.
PR 15-JAN-1999; 99US-0232880.
PR 09-APR-1999; 99US-0288946.
XX
PA (CORI-) CORIXA CORP.
XX
PI Dillon DC, Bardwaker SL, Yuqin J, Xu J, Mitcham JL;
XX
DR WPI; 2000-172129/15.
XX
PT New polypeptide useful for treating and diagnosing prostate cancer
PT comprises an immunogenic portion of prostate tumour protein -
XX
PS Claim 50; Page 239; 264pp; English.
XX
CC The present invention discloses isolated polypeptides, comprising an
CC immunogenic portion of a prostate tumour protein (PTP), the polypeptides
CC and polynucleotides encoding them have cytostatic activity and can be
CC used in vaccines and in gene therapy. The polypeptides and
CC polynucleotides encoding them, antibodies promoting cells which express
CC the polypeptides, antibodies against the polypeptides and vaccines
CC comprising them can be used for inhibiting the development of prostate
CC cancer in a patient. The polypeptides can be used to generate either for
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC the [illegible] describes that the polypeptides can be used as a probe or
CC to mediate the expression of the polypeptide. AAA06611 to AAA06691 and
CC AAY82707 to AAY82820 represent sequences used in the exemplification of
CC the present invention.
XX
SQ [illegible]; 49 A; 1 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_215_222 x AAA06617 ..

```
 Align seq 1/1  to: AAA06617  from: 1  to: 278

       1 MetLysGluArgThrSerProGly 8
         ||||||||||||||||||||||||
      92 ATGAAAGAGCGGACTTCACCTGGG 115

Seq name: /tmp/biodata/geneseq/geneseqp_NA_2001_SEARCH.seq

seq documentation block:
ID   AAF21828 standard; DNA; 1087 BP.
XX
AC   AAF21828;
XX
DT   27-MAR-2001  (first entry)
XX
DE   Human breast and ovarian cancer associated antigen gene SEQ ID 215.
XX
KW   Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
KW   nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW   antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
KW   antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
KW   Addison's disease; allergy; autoimmune haemolytic anaemia;
KW   autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW   multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW   cardiovascular disorder; wound healing; neurological disease; ds.
XX
OS   Homo sapiens.
XX
PN   WO200055173-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05881.
XX
PR   12-MAR-1999;  99US-0124270.
XX
PA   (HUMA ) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-611515/58.
DR   P-PSDB; AAB58925.
XX
PT   New human breast and ovarian cancer associated gene sequences and the
PT   polypeptides encoded by these genes, useful in the prevention,
PT   treatment and diagnosis of cancer, immune disorders, cardiovascular
PT   disorders and neurological diseases -
XX
PS   Claim 1; Page 642-643; 1299pp; English.
XX
CC   Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
CC   proteins AAB58711 - AAB59128. The DNA and protein sequences are
CC   associated with breast and ovarian cancer. Included in the invention are
CC   sequences AAF22042 - AAF22049 and AAB59129 which are used in the
CC   isolation and characterisation of the DNA and protein sequences of the
CC   invention. The breast and ovarian cancer associated DNA, protein, agonist
CC   or antagonist sequences exhibit cytostatic; immunosuppressive;
CC   nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
CC   antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
CC   antibacterial; antifungal; antiparasitic and cardiant activity. The
CC   polynucleotide and protein sequences are used in the diagnosis of cancer,
CC   particularly breast and ovarian cancer. The nucleic acid sequences,
CC   proteins, agonists and agonists may also be used in the diagnosis,
CC   prevention and treatment of immune disorders e.g. Addison's disease,
CC   allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC   diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC   arthritis and ulcerative colitis; cardiovascular disorders such as
CC   myocardial ischaemias; wound healing; neurological diseases such as
CC   cerebral anoxia and epilepsy; and infectious diseases.
XX
SQ   Sequence 1087 BP; 215 A; 347 C; 339 G; 182 T; 4 other;

alignment scores:
            Quality:    8.00              Length:       8
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 Seq name: /tmp/biodata/geneseq/geneseqp_NA_2001_SEARCH.seq vs AAF21828  ..

 Align seq 1/1  to: AAF21828  from: 1  to: 1087

       1 MetLysGluArgThrSerProGly 8
         ||||||||||||||||||||||||
     919 ATGAAAGAGCGGACTTCACCTGGG 942

Seq name: /tmp/biodata/geneseq/geneseqp_NA_2001_SEARCH.seq:AAZ50691

seq documentation block:
ID   AAZ50691 standard; cDNA; 1894 BP.
XX
AC   AAZ50691;
XX
DT   23-MAY-2000  (first entry)
XX
DE   Human prostate derived Ets factor cDNA.
XX
KW   Prostate derived Ets factor; PDEF; human; chromosome 6p21.3; cancer;
KW   loss of heterozygosity; chromosomal translocation; linkage analysis;
KW   cytostatic; cardiant; immunosuppressive; cerebroprotective; fungicide;
KW   antibacterial; vulnerary; neuroprotective; antiparkinsonian; nootropic;
KW   anabolic; antiinflammatory; anorectic; hybridisation probe; forensic;
KW   tumour marker; diagnosis; treatment; prostate cancer; blood coagulation;
KW   autoimmune disorder; haematopoietic; immune/nervous system; stroke;
KW   neoplasm; microbial infection; tissue regeneration; heart attack;
KW   scarring; food additive; preservative; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             416..1423
FT                   /*tag= a
FT                   /product= "Human prostate derived Ets factor"
FT                   /note= "Expressed in prostate epithelium"
XX
PN   WO200006589-A1.
XX
PD   10-FEB-2000.
XX
PF   02-AUG-1999;  99WO-US17470.
XX
PR   31-JUL-1998;  98US-0126945.
XX
PA   (HUMA ) HUMAN GENOME SCI INC.
PA   (BETH ) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI   Libermann IA,  Oettgen JP,  Kunsch CA,  Endress GA,  Rosen CA;
XX
DR   WPI; 2000-195255/17.
DR   P-PSDB; AAY45003.
XX
PT   Novel prostate derived polypeptide, polynucleotide useful for
PT   diagnosis, prevention and treatment of prostate cancer, autoimmune
PT   disorders, microbial infections and also as food additive or
PT   preservative
XX
PS   Claim 4; Fig 1; 142pp; English.
XX
CC   The present cDNA sequence encodes a human prostate derived Ets factor
CC   (PDEF), isolated from human prostate carcinoma cDNA library
CC   (ATCC No 203072). PDEF gene expression is highest in tissues involved in
CC   androgen sensitivity, with enriched expression in prostate epithelium
CC   cells. PDEF gene is mapped to the human chromosome 6p21.3
CC   region that is associated with loss of heterozygosity and chromosomal
CC   translocations in various human cancers. PDEF has cytostatic, cardiant,
```

```
CC   immunosuppressive, cerebroprotective, fungicide, antibacterial,
CC   vulnerary, neuroprotective, antiparkinsonian, nootropic, anatolic,
CC   antiinflammatory and anorectic activity. PDEF polynucleotides are useful
CC   in linkage analysis as markers, as hybridisation probes for differential
CC   identification of the tissues or cell types and as polymorphic markers
CC   for forensic purposes. PDEF is useful as prostate specific tumour marker
CC   for the diagnosis and treatment of prostate cancer. PDEF sequences are
CC   useful for treating autoimmune disorders, haematopoietic, blood
CC   coagulation, immune and nervous system disorders, hyperproliferative
CC   disorders like neoplasms and microbial infections, heart attacks
CC   stroke, scarring and for tissue regeneration. They are also useful as
CC   food additives or preservatives.
XX
SQ   Sequence 1894 BP; 368 A; 653 C; 571 G; 302 T; 0 other;


alignment_scores:
             Quality:    8.00              Length:       8
               Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_215_222 x AAZ50691    ..

 Align seq 1/1  to: AAZ50691  from: 1  to: 1894

        1 MetLysGluArgThrSerProGly 8
          ||||||||||||||||||||||||
     1058 ATGAAAGAGGGACTTCACCTGGG 1081

seq_name: gi|6537|gb|pda|pse seq p|pse:pp_PA2[1-1894] PA2 AAC84261

seq_documentation_block:
ID   AAC84261 standard; DNA; 1894 BP.
XX
AC   AAC84261;
XX
DT   16-MAR-2001  (first entry)
XX
DE   Gland-specific Ets transcription factor (GSEF) nucleotide sequence.
XX
KW   Transcription factor; gland-specific Ets transcription factor; GSEF;
KW   metastatic potential; cancer; tumour; metastasis; breast; prostate;
KW   leukaemia; lymphoma; sarcoma; melanoma; chromosome 6p21.1-6p21.3; ds.
XX
OS   Homo sapiens.
XX
PN   WO200070092-A1.
XX
PD   23-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US13173.
XX
PR   14-MAY-1999;   99US-0134112.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Kaufmann J,  Xin H,  Harrowe G;
XX
DR   WPI; 2001-041019/05.
DR   P-PSDB; AAB49628.
XX
PT   Detecting metastatic and potential metastatic cancerous cells, useful
PT   for diagnosing, prognosing, grading and staging of cancers by detecting
PT   gland-specific Ets transcription factor gene product in a biological
PT   sample from a cell -
XX
PS   Disclosure; Page 81-84; 95pp; English.
XX
CC   This invention relates to a method for the detection and determination of
CC   the metastatic potential of a cell. The method comprises detecting a
CC   gland-specific Ets transcription factor (GSEF) gene product in a test
CC   sample. Detection of a GSEF gene product in the test sample in amount
CC   lower than that in a normal cell, is indicative of a cell with high
CC   metastatic potential. The method is useful for determining the metastatic
CC   potential of a cell, for the diagnosis and prognosis of cancer as well as
CC   grading and staging of cancers by detecting GSEF expression in a
CC   biological test sample. The method may also be used to monitor patients
CC   having a predisposition to develop a particular cancer. GSEF polypeptides
CC   are useful for producing antibodies, in cancer diagnosis, prognosis,
CC   grading, staging and management of breast and prostate tumours, and in
CC   detecting polymorphisms in the sequence. GSEF genes and proteins are also
CC   useful in gene therapy. GSEF gene product expression levels can be used
CC   in conjunction with any tissue in which an alteration in GSEF gene
CC   product expression levels is associated with development of a
CC   cancer-associated phenotype. Cancers, which can be monitored include
CC   cancers of the prostate, cervix, lung and colon, melanomas, colorectal
CC   adenocarcinomas, Wilm's tumour, retinoblastoma, sarcomas, myosarcomas,
CC   lung carcinomas, leukaemia, and lymphomas. The GSEF gene is located on
CC   human chromosome 6, specifically at 6p21.1-6p21.3. The present sequence
CC   represents the DNA encoding GSEF
XX
SQ   Sequence 1894 BP; 368 A; 653 C; 571 G; 302 T; 0 other;


alignment_scores:
             Quality:    8.00              Length:       8
               Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_215_222 x AAC84261    ..

 Align seq 1/1  to: AAC84261  from: 1  to: 1894

        1 MetLysGluArgThrSerProGly 8
          ||||||||||||||||||||||||
     1058 ATGAAAGAGGGACTTCACCTGGG 1081

seq_name: gi|6537|gb|pda|pse seq p|pse:pp_PA2[1-1894] PA2 AAC84261

seq_documentation_block:
ID   AAC84261 standard; DNA; 1894 BP.
XX
AC   AAC84261;
XX
DT   16-MAR-2001  (first entry)
XX
DE   Gland-specific Ets transcription factor (GSEF) nucleotide sequence.
XX
KW   Transcription factor; gland-specific Ets transcription factor; GSEF;
KW   metastatic potential; cancer; tumour; metastasis; breast; prostate;
KW   leukaemia; lymphoma; sarcoma; melanoma; chromosome 6p21.1-6p21.3; ds.
XX
OS   Homo sapiens.
XX
PN   WO200070092-A1.
XX
PD   23-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US13173.
XX
PR   14-MAY-1999;   99US-0134112.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Kaufmann J,  Xin H,  Harrowe G;
XX
DR   WPI; 2001-041019/05.
DR   P-PSDB; AAB49633.
XX
PT   Detecting metastatic and potential metastatic cancerous cells, useful
PT   for diagnosing, prognosing, grading and staging of cancers by detecting
PT   gland-specific Ets transcription factor gene product in a biological
PT   sample from a cell
XX
```

PS Disclosure; Fig 2; 95pp; English.
XX
CC This invention relates to a method for the detection and determination of
CC the metastatic potential of a cell. The method comprises detecting a
CC gland specific Ets transcription factor (GSEF) gene product in a test
CC sample. Detection of a GSEF gene product in the test sample in amount
CC lower than that in a normal cell, is indicative of a cell with high
CC metastatic potential. The method is useful for determining the metastatic
CC potential of a cell, for the diagnosis and prognosis of cancer as well as
CC grading and staging of cancers by detecting GSEF expression in a
CC biological test sample. The method may also be used to monitor patients
CC having a predisposition to develop a particular cancer. GSEF polypeptides
CC are useful for producing antibodies, in cancer diagnosis, prognosis,
CC grading, staging and management of breast and prostate tumours, and in
CC detecting polymorphisms in the sequence. GSEF genes and proteins are also
CC useful in gene therapy. GSEF gene product expression levels can be used
CC in conjunction with any tissue in which an alteration in GSEF gene
CC product expression levels is associated with development of a
CC cancer associated phenotype. Cancers which can be monitored include
CC cancers of the prostate, cervix, lung and colon, melanomas, colorectal
CC adenocarcinomas, Wilm's tumour, retinoblastoma, sarcomas, myosarcomas,
CC lung carcinomas, leukaemia, and lymphomas. The GSEF gene is located on
CC human chromosome 6, specifically at 6p21.1-6p21.3. The present sequence
CC represents the cDNA encoding GSEF.
XX
SQ Sequence 3317 BP; 710 A; 1026 C; 970 G; 611 T; 0 other;

alignment scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2_COPY_215_222 x AAC83266 ..

Align seq 1/1 to: AAC83266 from: 1 to: 3317

1 MetLysGlnArgIleSerProGly 8
||||||||||||||||||||||||
2457 ATGAAACAGCGGATTTCACCTGGG 2480

seq name: /cgbin/gcg-pdf-dir/gcg.seq-locus-no-NAI-1999-Apr-AAV89435

seq documentation block:
ID AAV89435 standard; cDNA; 248 BP.
XX
AC AAV89435;
XX
DT 15-FEB-1999 (first entry)
XX
DE EST clone C0851.
XX
KW Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
KW tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;
KW receptor; ligand; thrombolytic; anti inflammatory; cadherin; anti-tumour;
KW gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO9845446-A2.
XX
PD 15-OCT-1998.
XX
PF 10-APR-1998; 98WO-US06955.
XX
PR 10-APR-1997; 97US-0838821
XX
PA (GEMY) GENETICS INST INC.
XX
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI Racie LA, Spaulding V, Treacy M;
XX
DR WPI; 1999-020077/02.
XX
PT New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries.
XX
PS Claim 1; Page 217; 618pp; English.
XX
CC The present sequence represents a human expressed sequence tag (EST).
CC The polynucleotide, which is a secreted EST, and the encoded protein
CC are predicted to have useful biological activities which would make
CC them suitable for treating, preventing or ameliorating medical
CC conditions in humans and animals, although no supporting data is
CC given. Suggested activities include nutritional activity, immune
CC stimulating or suppressing activity, haematopoiesis regulating
CC activity, tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC activity, receptor/ligand activity, anti inflammatory activity,
CC cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The polynucleotide may also be useful for gene therapy.
XX
SQ Sequence 248 BP; 65 A; 56 C; 62 G; 65 T; 0 other;

alignment scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2_COPY_215_222 x AAV89435 ..

Align seq 1/1 to: AAV89435 from: 1 to: 248

8 GluArgIleSerProGly 8
|||||||||||||||||||
209 GAGCGAACCTCCCCAGGC 226

seq name: /cgbin/gcg-pdf-dir/gcg.seq-locus-no-NAI-1999-Apr-AAQ61310

seq documentation block:
ID AAQ61310 standard; DNA; 276 BP.
XX
AC AAQ61310;
XX
DT 16-MAR-1994 (first entry)
XX
DE Human brain Expressed Sequence Tag EST01419.
XX
KW Gene transcription product; genetic markers; tagging; in vivo;
KW transcription; mapping; locations; chromosomes; chromosomal; ss.
XX
OS Homo sapiens.
XX
PN WO9416178-A.
XX
PD 19-AUG-1994.
XX
PF 12-FEB-1993; 93WO-US01419.
XX
PR 12-FEB-1992; 92US-0834195.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
XX
PI Adams MD, Moreno RF, Venter CJ;
XX
DR WPI; 1994-272882/34.
XX
PT Enriched oligonucleotides and corresp. sequences - used as
PT markers for human genes transcribed in vivo, facilitate tagging
PT of most human genes.
XX
PS Example 4; Page 471; 500pp; English.

```
XX
CC   The Expressed Sequence Tag was isolated from a human brain cDNA
CC   library as part of a large set of ESTs which can be used as markers
CC   for human genes transcribed in vivo. They can be used to facilitate
CC   tagging of most human genes, for mapping locations of expressed genes
CC   on chromosomes, for individual or forensic identification, for mapping
CC   locations of disease-associated genes, for identification of tissue
CC   type, and for prepn. of antisense sequences, probes and constructs.
CC   EST01319 has a "poor" coding probability as evaluated using the
CC   coding-region prediction program CRM. See also AAQ59041-Q61440.
XX
SQ   Sequence 276 BP; 66 A; 68 C; 86 G; 54 T; 2 other;


alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000              Gaps:         0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_215_222 x AAQ61310    ..

 Align seq 1/1  to: AAQ61310  from: 1  to: 276

        3 GluArgThrSerProGly 8
          ||||||||||||||||||
      167 GAAAGAACTTCCCCAGGA 184

seq_name: ...; ...; data_pro seq_p ...p_NA; ...; AA...

seq_documentation_block:
ID   AAC07838 standard; cDNA; 297 BP.
XX
AC   AAC07838;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID No: 11914.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-509481/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures
XX
PS   Claim 1; SEQ ID 11913; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 297 BP; 92 A; 44 C; 62 G; 98 T; 2 other;


alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000              Gaps:         0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_215_222 x AAC07838    ..

 Align seq 1/1  to: AAC07838  from: 1  to: 297

        2 LysGluArgThrSerPro 7
          || |||||||||||||||
      217 AAGGAAGAACTTCCCCAGGA 236

seq_name: ...; ...; data_pro seq_p ...p_NA; ...; AA...

seq_documentation_block:
ID   AAZ14159 standard; cDNA; 600 BP.
XX
AC   AAZ14159;
XX
DT   12-OCT-1999  (first entry)
XX
DE   Human gene expression product cDNA sequence SEQ ID NO:628.
XX
KW   Human; gene; gene expression product; diagnosis; therapy; probe;
KW   detection; mapping; tissue typing; profiling; forensic; cancer;
KW   genetic analysis; colorectal cancer, breast cancer, lung cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO9938972-A2.
XX
PD   05-AUG-1999.
XX
PF   28-JAN-1999;   99WO-US01619.
XX
PR   03-APR-1998;   98US-0080666.
PR   28-JAN-1998;   98US-0072910.
PR   24-FEB-1998;   98US-0075954.
PR   31-MAR-1998;   98US-0080114.
PR   03-APR-1998;   98US-0080515.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE ) HYSEQ INC.
XX
PI   Cirkovenkov K,  Dickson M,  Drmanac R,  Drmanac S;
PI   Escobedo J,  Garcia PD,  Garcia V,  Glese K,  Innis MA;
PI   Jones WL,  Kassam A,  Kennedy GC,  Kita D,  Labat I;
PI   Lamson G,  Loshkowitz D,  Pau B,  Pandeggo F,  Reinhard C;
PI   Stuelpnagin B,  Suddath Klinger J,  Williams LE;
XX
DR   WPI; 1999-494092-41.
XX
PT   Novel human genes and their expression products which are
PT   differentially expressed in different cell types
XX
PS   Claim 1; Page 805; 247pp; English.
XX
CC   The present invention describes a library of human polynucleotides
CC   comprising the sequences given in AAZ12542 to AAZ17779. Also described is
CC   a method of detecting differentially expressed genes correlated with the
CC   cancerous state of a mammalian cell, comprising detecting at least one
CC   differentially expressed gene product in a test sample from a cell
```

CC suspected of being cancerous, where the gene product is encoded by one
CC of the 5248 polynucleotide sequences given in AA212562 to AAZ17779. The
CC polynucleotides can be used as a source of primers and probes, which can
CC be used for a variety of purpose, e.g. detection of expression levels,
CC mapping, tissue typing or profiling, forensics, genetic analysis and
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists.
XX
SQ Sequence 300 BP; 85 A; 70 C; 66 G; 79 T; 0 other;

alignment scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2-COPY-215_222 x AAZ13159 ..

Align Seq 1/1 to: AAZ13159 from: 1 to: 300

3 GluArgThrSerProGly 8
277 GAGCGAACCTCCCAGGC 294

seq name: /SIDS1/seqdata/geneseq/geneseqp/NA2000.DAT:AAZ13159

seq documentation block:
ID AAZ13159 standard; cDNA; 300 BP.
XX
AC AAZ13159;
XX
DT 19-MAY-2000 (first entry)
XX
DE Human colon cancer cell line polynucleotide sequence SEQ ID NO:264.
XX
KW Human; colon cancer; tumour; diagnosis; gene expression product;
KW probe; detection; cancerous state; metastasis; identification;
KW breast cancer; oestrogen receptor-positive breast cancer; therapy;
KW oestrogen receptor-negative breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9958675-A2.
XX
PD 18-NOV-1999.
XX
PF 13-MAY-1999; 99WO-US10602.
XX
PR 14-MAY-1998; 98US-0085426.
PR 15-MAY-1998; 98US-0085537.
PR 15-MAY-1998; 98US-0085696.
PR 21-OCT-1998; 98US-0105234.
PR 27-OCT-1998; 98US-0105877.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Suddath Klinger J;
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache Crain B;
XX
DR WPI; 2000-126469/11.

XX
PT Polynucleotide library, used to determine cancerous states of mammalian
PT cells
XX
PS Claim 1; Page 243; 1097pp; English.
XX
CC AAA00016 to AAA02716 represent polynucleotides isolated from cDNA
CC libraries constructed from human colon cancer cell lines. The present
CC invention also describes a method of detecting differentially expressed
CC genes correlated with a cancerous state of a mammalian cell, comprising
CC detecting at least one differentially expressed gene product in a test
CC sample derived from a cell suspected of being cancerous, where detection
CC of the differentially expressed gene product is correlated with a
CC cancerous state of the cell from which the test sample was derived.
CC The polynucleotides sequences can be used in a method for detecting
CC differentially expressed genes correlated with a cancerous state of a
CC mammalian cell. The polynucleotides can also be used as probes for
CC detecting and mapping related genes. They can be used in diagnosis and
CC prognosis of diseases and disorders (e.g. identification of
CC pre-metastatic or metastatic cancerous states, stages of cancer, or
CC responsiveness of cancer to therapy). This is particularly for breast
CC cancer, oestrogen receptor positive breast cancer, oestrogen receptor
CC negative breast cancer, lung cancer, and colon cancer.
XX
SQ Sequence 300 BP; 91 A; 72 C; 92 G; 45 T; 0 other;

alignment scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2-COPY-215_222 x AAA00273 ..

Align Seq 1/1 to: AAA00273 from: 1 to: 300

3 GluArgThrSerProGly 8
45 GAAAGAACCAGCCAGGG 62

seq_name: /SIDS1/seqdata/geneseq/geneseqp/NA2000.DAT:AAA00351

seq_documentation_block:
ID AAA00351 standard; cDNA; 300 BP.
XX
AC AAA00351;
XX
DT 19-MAY-2000 (first entry)
XX
DE Human colon cancer cell line polynucleotide sequence SEQ ID NO:342.
XX
KW Human; colon cancer; tumour; diagnosis; gene expression product;
KW probe; detection; cancerous state; metastasis; identification;
KW breast cancer; oestrogen receptor positive breast cancer; therapy;
KW oestrogen receptor negative breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9958675-A2.
XX
PD 18-NOV-1999.
XX
PF 13-MAY-1999; 99WO-US10602.
XX
PR 14-MAY-1998; 98US-0085426.
PR 15-MAY-1998; 98US-0085537.
PR 15-MAY-1998; 98US-0085696.
PR 21-OCT-1998; 98US-0105234.
PR 27-OCT-1998; 98US-0105877.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE) HYSEQ INC.

```
XX
PI   Williams LI,  Escobedo J,  Innis MA,  Garcia PD,  Suddath Klinger J;
PI   Reinhard C,  Giese K,  Randazzo F,  Kennedy G,  Fot D,  Kassam A;
PI   Lamson G,  Drmanac R,  Crkvenjakov R,  Dickson M,  Drmanac S,  Lahat I;
PI   Leshkowitz D,  Kita D,  Garcia V,  Jones LW,  Stache-Crain B;
XX
DR   WPI; 2000-126469/11.
XX
TI   Polynucleotide library used to determine cancerous states of mammalian
TI   cells
XX
PS   Claim 1; Page 262; 1097pp; English.
XX
CC   AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
CC   libraries constructed from human colon cancer cell lines. The present
CC   invention also describes a method of detecting differentially expressed
CC   genes correlated with a cancerous state of a mammalian cell, comprising
CC   detecting at least one differentially expressed gene product in a test
CC   sample derived from a cell suspected of being cancerous, where detection
CC   of the differentially expressed gene product is correlated with a
CC   cancerous state of the cell from which the test sample was derived.
CC   The polynucleotides sequences can be used in a method for detecting
CC   differentially expressed genes correlated with a cancerous state of a
CC   mammalian cell. The polynucleotides can also be used as probes for
CC   detecting and mapping related genes. They can be used in diagnosis and
CC   prognosis of diseases and disorders (e.g. identification of
CC   pre-metastatic or metastatic cancerous states, stages of cancer, or
CC   responsiveness of cancer to therapy). This is particularly for breast
CC   cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
CC   negative breast cancer, lung cancer, and colon cancer.
XX
SQ   Sequence 300 BP; 94 A; 53 C; 74 G; 79 T; 0 other;


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_215_222 x AAA00351   ..

 Align seq 1/1  to: AAA00351  from: 1  to: 300

       2 LysGluArgThrSerPro 7
         |||||||||||||||||
     277 AAGGAAAGAACTCCTCCA 294

seq_name: /cds1/geninfo/genseqs/genseq_NA2000.DAT:AAA00643

seq_documentation_block:
ID   AAA00643 standard; cDNA; 300 BP.
XX
AC   AAA00643;
XX
DT   19-MAY-2000  (first entry)
XX
DE   Human colon cancer cell line polynucleotide sequence SEQ ID NO:634.
XX
KW   Human; colon cancer; tumour; diagnosis; gene expression product;
KW   probe; detection; cancerous state; metastasis; identification;
KW   breast cancer; oestrogen receptor positive breast cancer; therapy;
KW   oestrogen receptor-negative breast cancer; lung cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO9958675-A2.
XX
PD   18-NOV-1999.
XX
PF   13-MAY-1999;   99WO-US10602.
XX
PR   14-MAY-1998;   98US-0085426.
PR   15-MAY-1998;   98US-0085537.
PR   15-MAY-1998;   98US-0085696.
PR   21-OCT-1998;   98US-0105234.
PR   27-OCT-1998;   98US-0105877.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE ) HYSEQ INC.
XX
PI   Williams LI,  Escobedo J,  Innis MA,  Garcia PD,  Suddath-Klinger J;
PI   Reinhard C,  Giese K,  Randazzo F,  Kennedy G,  Fot D,  Kassam A;
PI   Lamson G,  Drmanac R,  Crkvenjakov R,  Dickson M,  Drmanac S,  Lahat I;
PI   Leshkowitz D,  Kita D,  Garcia V,  Jones LW,  Stache-Crain B;
XX
DR   WPI; 2000-126469/11.
XX
TI   Polynucleotide library used to determine cancerous states of mammalian
TI   cells
XX
PS   Claim 1; Page 332; 1097pp; English.
XX
CC   AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
CC   libraries constructed from human colon cancer cell lines. The present
CC   invention also describes a method of detecting differentially expressed
CC   genes correlated with a cancerous state of a mammalian cell, comprising
CC   detecting at least one differentially expressed gene product in a test
CC   sample derived from a cell suspected of being cancerous, where detection
CC   of the differentially expressed gene product is correlated with a
CC   cancerous state of the cell from which the test sample was derived.
CC   The polynucleotides sequences can be used in a method for detecting
CC   differentially expressed genes correlated with a cancerous state of a
CC   mammalian cell. The polynucleotides can also be used as probes for
CC   detecting and mapping related genes. They can be used in diagnosis and
CC   prognosis of diseases and disorders (e.g. identification of
CC   pre-metastatic or metastatic cancerous states, stages of cancer, or
CC   responsiveness of cancer to therapy). This is particularly for breast
CC   cancer, oestrogen receptor positive breast cancer, oestrogen receptor
CC   negative breast cancer, lung cancer, and colon cancer.
XX
SQ   Sequence 300 BP; 80 A; 55 C; 77 G; 88 T; 0 other;


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_215_222 x AAA00643/rev   ..

 Align seq 1/1  to reverse of: AAA00643  from: 1  to: 300

       3 GluArgThrSerProGly 8
         ||||:||||||||||||
      93 GAGAGAACAAGTCCTGGA 76

seq_name: /cds1/geninfo/genseqs/genseq_NA1992.DAT:AAQ20365

seq_documentation_block:
ID   AAQ20365 standard; DNA; 312 BP.
XX
AC   AAQ20365;
XX
DT   09-APR-1992  (first entry)
XX
DE   Probe specific to the gp48 gene region of BVDV genome.
XX
KW   Polymerase chain reaction; PCR; DNA amplification; cattle; ss.
XX
OS   Synthetic.
XX
PN   EP464010-A.
XX
PD   02-JAN-1992.
```

XX
PF 24 MAY 1991; 91EP 0850139.
XX
PR 08 JUN 1990; 90SE 0002060.
XX
PA (STAT) STATENS VETERINARME.
XX
PI Belak S, Belloqi Pordany A;
XX
DR WPI; 1992 009568/02.
XX
PI Detection of bovine viral diarrhoea virus infection using
PT primers and probes from the qp48 nucleotide region of 1362 1880
PT of bovine viral diarrhoea virus genome
XX
PS Claim 9; Page 14; 17pp; English.
XX
CC BVDV DNA amplified using primers corresponding to sequences located
CC within the qp48 nucleotide region of 1362 1880 of the BVDV genome
CC is detected using this probe. See also AAQ20361-4.
XX
SQ Sequence 312 BP; 97 A; 68 C; 76 G; 71 T; 0 other;

alignment scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2/copy_215_222 x AAQ20365 ..

Align seq 1/1 to: AAQ20365 from: 1 to: 312

2 LysGluArgThrSerPro 7
||||||||||||||||||
143 AAGGAAAGAACTTCTCCT 160

seq name: /cgbsl/geqdata/geneseq/geneseqp_NA20_3.1/A/AAQ20365

seq documentation block:
ID AAQ20365 standard; DNA; 312 BP.
XX
AC AAQ20365;
XX
DT 02 FEB 2001 (first entry)
XX
DE Pollen allergen promoter coding sequence #1.
XX
KW Promoter; eucalyptus; pine; gene transcription; ds.
XX
OS Eucalyptus grandis
XX
PN WO200058474 A1.
XX
PD 05 OCT 2000.
XX
PF 24 FEB 2000; 2000WO NZ00018.
XX
PR 25 MAR 1999; 99US-0276599.
PR 30 JUL 1999; 99US-0146591.
XX
PA (GENE) GENESIS RES & DEV CORP LTD.
PA (FLET) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Perera R, Rice SJ, Eagleton CK;
XX
DR WPI; 2000-647236/62
XX
PT Novel promoter sequences useful for modulating transcription of plant
PT DNA sequences of interest and production of polypeptides
XX
PS Claim 1; Page 51; 93pp; English.

XX
CC The present invention relates to promoter sequences from eucalyptus
CC and pine. The present sequence is one such promoter. This sequence is
CC useful for modulating the transcription of DNA sequences of interest. The
CC sequences may also be used to tag or identify an organism or its
CC reproductive material.
XX
SQ Sequence 350 BP; 119 A; 59 C; 51 G; 121 T; 0 other;

alignment scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 120.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2/copy_215_222 x AAC62773/rev ..

Align seq 1/1 to reverse of: AAC62773 from: 1 to: 350

3 GluArgThrSerProGly 8
||||||||||||||||||
41 GAAAGGACCAGCCCGGGC 24

Seq name: /cgbsl/geqdata/geneseq/geneseqp_NA2000.dAt:AAC30140

seq documentation block:
ID AAC30140 standard; cDNA; 350 BP.
XX
AC AAC30140;
XX
DT 06 OCT 2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 34215.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401 A2.
XX
PD 06-SEP 2000.
XX
PF 21 FEB 2000; 2000EP 0200610.
XX
PR 26 FEB 1999; 99US-0122487.
XX
PA (GESI) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures
XX
PS Claim 1; SEQ ID 34215; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design

```
CC   expression and secretion vectors.
XX
SQ   Sequence 350 BP; 120 A; 55 C; 82 G; 92 T; 1 other;


Alignment_scores:
            Quality:    6.00              Length:       5
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment_block:
 US-09-126-945B-2_COPY_215_222 x AAC30140  ..

 Align seq 1/1  to: AAC30140  from: 1  to: 350

       2 LysGluArgThrSerPro 7
         |||||||||||||||||
       9 AAGGAAAGAACTTCTCCA 26
```

```
OM of: US-09-126-945B-2_COPY_215_222 fro[illegible]

Date: Sep 14, 2001  9:53 PM

About: Results were produced by the [illegible]
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
  -MODEL frame+_p2n.model -DEV=xlp
  -Q "[illegible]" -p="[illegible]"
  -DB Issued_Patents_NA -QFMT=fastap -SUFFIX olip2n.rni
  -GAPOP 4.500 -GAPEXT=0.050 -MINMATCH=0.100 -LoopCL 0.000
  -LOOPEXT 0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP 60.000
  -XGAPEXT 60.000 -FGAPOP=6.000 -FGAPEXT 7.000 -YGAPOP 60.000
  -YGAPEXT 60.000 -DFLOP=6.000 -DELEXT=7.000 -START 1
  -MATRIX olig5 -TRANS=human40.cdi -LIST 45 -DOCALIGN 200
  -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT pir
  -NORM ext  MINLEN 0 -MAXLEN 2000000000
  -USER=US0126945_9CON1_1_91 -NCPU=6 -JCPU 3 -JLNCL=3 -NO_XLPXY
  -WAIT -THREADS=1

Search Information block:
Query: US-09-126-945B-2_COPY_215_222
Query length: 8
Database: Issued_Patents_NA:*
Database sequences: 424599
Database length: 94655562
Search time (sec): 322.720000

WARN: XGAPOP and YGAPOP must be equal, Assuming YGAPOP XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal, Assuming YGAPEXT XGAPEXT 60.000


score_list:
Sequence          Strd Orig    ZScore   EScore Len   ! Documentation ..
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-670-964-1 +     6.00  102.21  416.29     1750  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-670-964-1 +     6.00  102.20  416.42     1987  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-[illegible]
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-670-[illegible]
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-923-151A-4 +     6.00  102.08  403.97     2764  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-888-[illegible]
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-706-[illegible]
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-706-4410-29 +     6.00  102.08  404.10     2765  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-706-4410-29 +     6.00  102.08  404.10     2765  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-706-4410-[illegible]
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-967-101-1 +     6.00  102.92  407.54     2791  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-967-101-1A6 +     6.00  102.92  407.54     2791  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-592-541-1 +     6.00  102.92  407.54     2791  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-592-541-1A6 +     6.00  102.92  407.54     2791  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-124-698-1 +     6.00  102.92  407.54     2791  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-124-698-1A6 +     6.00  102.92  407.54     2791  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-127-480-1 +     6.00  102.92  407.54     2791  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-127-480-1A6 +     6.00  102.92  407.54     2791  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-496-[illegible]
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-888-077A-4 +     6.00  102.21  416.29     3086  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-967-101-5 +     6.00  102.20  416.42     2987  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-592-541-5 +     6.00  102.20  416.42     2987  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-124-698-5 +     6.00  102.20  416.42     2987  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-127-480-5 +     6.00  102.20  416.42     2987  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-496-[illegible]
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-924-[illegible]
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-051-[illegible]
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-[illegible]
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-854-[illegible]
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-585-[illegible]
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-[illegible]
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-164-[illegible]
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-[illegible]
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-07-[illegible]
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-[illegible]
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-[illegible]
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-160-142-40 +     5.00  116.94  67.41     26  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-[illegible]
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-907-906-940E-15 +     5.00  116.42  72.08     28
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-[illegible]
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-908-941-26 +     5.00  116.94  76.72     30  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-[illegible]
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-102-4208-29 +     5.00  116.94  76.72     30

>1_1 Sequence 6 /ptodata/2/ina/6A_COMB.seq:US-08-670-964-3

seq_documentation_block:
; Sequence 3, Application US/08670964
; Patent No. 6010874
;  GENERAL INFORMATION:
;     APPLICANT:  Hardy, John A.
;     TITLE OF INVENTION:  EARLY ONSET ALZHEIMER'S DISEASE
;     TITLE OF INVENTION:  GENE AND GENE PRODUCTS
;     NUMBER OF SEQUENCES:  4
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  SmithKline Beecham Corporation
;       STREET:  709 Swedeland Road - UW2220; P.O. Box 15
;       CITY:  Philadelphia
;       STATE:  PA
;       COUNTRY:  USA
;       ZIP:  19406
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/670,964
;       FILING DATE:  26-JUN-1996
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  60/001,112
;       FILING DATE:  14-JUL-1995
;       APPLICATION NUMBER:  60/001,501
;       FILING DATE:  18-JUL-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Han, William T
;       REGISTRATION NUMBER:  34,844
;       REFERENCE/DOCKET NUMBER:  PS0458
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  610-270-5219
;       TELEFAX:  610-270-5090
;       TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1750 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-670-964-3


alignment_scores:
              Quality:    6.00              Length:      6
                 Ratio:  1.000                 Gaps:      0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_215_222 x US-08-670-964-3  ..

 Align_set 1,1  1..6  to  US-08-670-964-3  from: 1  to: 1750

        1 Matching_CDNAregion_Ser 6
          ||||||||||||||||||
      121 ACCGAAGCAAAGAAAGTCCA 127


>1_1 /ptodata/2/ina/6A_COMB.seq:US-08-670-964-1
```

```
Seq documentation block:
; Sequence 1, Application US/08/670964
; Patent No. 6010874
;  GENERAL INFORMATION:
;     APPLICANT:  Hardy, John A.
;     TITLE OF INVENTION:  EARLY ONSET ALZHEIMER'S DISEASE
;     TITLE OF INVENTION:  GENE AND GENE PRODUCTS
;     NUMBER OF SEQUENCES:  4
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  SmithKline Beecham Corporation
;       STREET:  709 Swedeland Road  UW2220; P.O. Box 15
;       CITY:  Philadelphia
;       STATE:  PA
;       COUNTRY:  USA
;       ZIP:  19406
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/670,964
;       FILING DATE:  26-JUN 1996
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  60/001,142
;       FILING DATE:  13 JUL 1995
;       APPLICATION NUMBER:  60/001,501
;       FILING DATE:  18 JUL 1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Han, William T
;       REGISTRATION NUMBER:  34,344
;       REFERENCE/DOCKET NUMBER:  P50458
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  610-270 5219
;       TELEFAX:  610 270-5090
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1762 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-670-964-1


Alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-126-945b-2_copy_215_222 x US-08-670-964-1  ..

 Align seq 1/1  to: US-08-670-964-1  from: 1  to: 1762

        1 MetLysGluArgThrSer 6
          ||||||||||||||||||
      110 ATGAAAGAAAGAACTCA 127

Seq name: /tmp2/pfr-data/215_222_copy_comMb.seq:US-08-909-965C-17

Seq documentation block:
; Sequence 17, Application US/08/909,965C
; Patent No. 5936078
;  GENERAL INFORMATION:
;     APPLICANT:  Keja Petsch
;     APPLICANT:  Nakagawa Satoshi
;     APPLICANT:  Sakaki yoshiyuki
;     APPLICANT:  Zhao Nanding
;     APPLICANT:  Hashida Hideji
;     TITLE OF INVENTION:  NOVEL DNA, NOVEL POLYPEPTIDE
;     TITLE OF INVENTION:  AND NOVEL ANTIBODY
;     NUMBER OF SEQUENCES:  17
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  FITZPATRICK, CELLA, HARPER AND SCINTO
;       STREET:  277 Park Avenue
;       CITY:  New York
;       STATE:  New York
;       COUNTRY:  U.S.A.
;       ZIP:  10172-0194
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/909,965C
;       FILING DATE:  August 12, 1997
;       CLASSIFICATION:  514
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  JP 322745/95
;       APPLICATION NUMBER:  PCT/JP96/03630
;       FILING DATE:  12-No. 5936078-1995
;       FILING DATE:  12-Dec-1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Lawrence S. Perry
;       REGISTRATION NUMBER:  31865
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  212-758-2400
;       TELEFAX:  212 758-2982
;       TELEX:  236262
;   INFORMATION FOR SEQ ID NO:  17:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1908 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA to mRNA
;     ORIGINAL SOURCE:
;       ORGANISM:  human\
;     IMMEDIATE SOURCE:
;       CLONE:  PZ593
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  640 to 1002
;       IDENTIFICATION METHOD:  by experiment
US-08-909-965C-17


Alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-126-945b-2_copy_215_222 x US-08-909-965C-17  ..

 Align seq 1/1  to: US-08-909-965C-17  from: 1  to: 1908

        8 GluArgThrSerProGly 8
          ||||||||||||||||||
     1103 GAGAAGACATCTCCTGGG 1120

Seq name: /tmp2/pfr-data/215_222_copy_comMb.seq:US-08-670-479-24

Seq documentation block:
; Sequence 24, Application US/08670479
; Patent No. 5723333
;  GENERAL INFORMATION:
;     APPLICANT:  Hardy, John A.
;     APPLICANT:  Goate, Alison M.
;     TITLE OF INVENTION:  MUTANT SIR2 GENES
;     NUMBER OF SEQUENCES:  24
;     CORRESPONDENCE ADDRESS:
```

```
;       ADDRESSEE:  SmithKline Beecham Corporation
;       STREET:  709 Swedeland Road
;       CITY:  King of Prussia
;       STATE:  PA
;       COUNTRY:  U.S.A.
;       ZIP:  19406-0939
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/670,479
;       FILING DATE:  26-JUN-1996
;       CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  60/001,500
;       FILING DATE:  18 JUL 1996
;       APPLICATION NUMBER:  60/001,800
;       FILING DATE:  02-AUG-1995
;    ATTORNEY/AGENT INFORMATION:
;       NAME:  Han, William T
;       REGISTRATION NUMBER:  34,444
;       REFERENCE/DOCKET NUMBER:  P50361
;    TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  610-270-5219
;       TELEFAX:  610-270-5090
;       TELEX:
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH:  1914 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
;    FEATURE:
US-08-670-479-24


alignment_scores:
            Quality:     6.00              Length:       6
              Ratio:    1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_215_222 x US-08-670-479-24    ..

 Align seq 1/1   to: US-08-670-479-24   from: 1   to: 1914

        1 MetTyrGlnArgThrSer 6
          ||||||||||||||||||
      156 AtGAAAGAAAGAACCtCA 173

seq_name: /cgn2_6/ptodata/2/iia/6A_COMB.seq;US-08-923-454A-9

seq_documentation_block:
; Sequence 9, Application US/08923454A
; Patent No. 6004794
;  GENERAL INFORMATION:
;     APPLICANT:  Creasy, Caretha
;     APPLICANT:  Livi, George
;     APPLICANT:  Karran, Eric
;     APPLICANT:  Clinkenbeard, Helen
;     APPLICANT:  Browne, Michael
;     APPLICANT:  Southan, Christopher
;     TITLE OF INVENTION:  HUMAN SERINE PROTEASE
;     NUMBER OF SEQUENCES:  40
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  SmithKline Beecham Corporation
;       STREET:  709 Swedeland Road
;       CITY:  King of Prussia
;       STATE:  PA
;       COUNTRY:  USA
;       ZIP:  19406
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ Version 1.5
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/923,454A
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  60/025436
;       FILING DATE:  06-SEPT-1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Barmeister, Kirk
;       REGISTRATION NUMBER:  33,833
;       REFERENCE/DOCKET NUMBER:  P50347
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  610-270-5096
;       TELEFAX:  610-270-5090
;       TELEX:
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH:  2764 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
US-08-923-454A-9


alignment_scores:
            Quality:     6.00              Length:       6
              Ratio:    1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_215_222 x US-08-923-454A-9    ..

 Align seq 1/1   to: US-08-923-454A-9   from: 1   to: 2764

        1 MetTyrGlnArgThrSer 6
          ||||||||||||||||||
      185 AtGAAAGAAAGAACCtCA 202

seq_name: /cgn2_6/ptodata/2/iia/6A_COMB.seq;US-08-923-454A-9

seq_documentation_block:
; Sequence 1, Application US/08888077A
; Patent No. 6020143
;  GENERAL INFORMATION:
;     APPLICANT:  ST. GEORGE-HYSLOP, PETER H
;     APPLICANT:  ROMMENS, JOHANNA M
;     APPLICANT:  FRASER, PAUL E
;     TITLE OF INVENTION:  GENETIC SEQUENCES AND PROTEINS RELATED
;     TITLE OF INVENTION:  TO ALZHEIMER'S DISEASE AND USES THEREFOR
;     NUMBER OF SEQUENCES:  41
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  LERNER, DAVID, LITTENBERG, KRUMHOLZ & MENTLIK
;       STREET:  600 SOUTH AVENUE WEST
;       CITY:  WESTFIELD
;       STATE:  NJ
;       COUNTRY:  USA
;       ZIP:  07090-1497
;     COMPUTER READABLE FORM:
```

```
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS DOS
;       SOFTWARE:  ASCII
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/888,077A
;       FILING DATE:  04-JUL-1997
;       CLASSIFICATION:  530
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/592,541
;       FILING DATE:  26-JAN-1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  PALIST, THOMAS M.
;       REGISTRATION NUMBER:  36,629
;       REFERENCE/DOCKET NUMBER:  SCHERING 3.0-017-CIP-CIP-CIP-IV
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (908) 654-5000
;       TELEFAX:  (908) 654-7866
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2765 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  249..1649
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  1..2675
;       OTHER INFORMATION:  /note= "UPST 1"
US 08 888 077A 1


alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US 09 126 945B 2 COPY 215 222 x US 08 888 077A 1   ..

 Align seq 1/1  to: US 08 888 077A 1  from: 1  to: 2765

        1 MetLysGluArgThrSer 6
          ||||||||||||||||||
      185 ATGAAAGAAAGAACCTCA 202

seq name: /cgn2_6/ptodata/2/lca/6B_COMP/seq.US 08 706 344C 1

seq documentation block:
; Sequence 1, Application US/08706344C
; Patent No. 6248555
;  GENERAL INFORMATION:
;     APPLICANT:  TANZI, RUDOLPH
;     APPLICANT:  WASCO, WILMA
;     TITLE OF INVENTION:  Genetic Alterations Related To Familial
;     TITLE OF INVENTION:  Alzheimer's Disease
;     NUMBER OF SEQUENCES:  42
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
;       STREET:  1100 NEW YORK AVENUE, SUITE 600
;       CITY:  WASHINGTON
;       STATE:  DC
;       COUNTRY:  USA
;       ZIP:  20005-3934
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/706,344C
;       FILING DATE:  30-AUG-1996
;       CLASSIFICATION:  536
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  60/003,054
;       FILING DATE:  31-AUG-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  KIM, JUDITH U.
;       REGISTRATION NUMBER:  40,679
;       REFERENCE/DOCKET NUMBER:  0609.41B0001
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  202-371-2600
;       TELEFAX:  202-371-2540
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2765 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  249..1649
US 08 706-344C 1


alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US 09-126 945B 2 COPY_215 222 x US 08 706 344C 1   ..

 Align seq 1/1  to: US 08 706 344C 1  from: 1  to: 2765

        1 MetLysGluArgThrSer 6
          ||||||||||||||||||
      185 ATGAAAGAAAGAACCTCA 202

seq name: /cgn2_6/ptodata/2/lca/6B_COMP/seq.US 08 706 344C 1

seq documentation block:
; Sequence 1, Application US/08706344C
; Patent No. 6248555
;  GENERAL INFORMATION:
;     APPLICANT:  TANZI, RUDOLPH
;     APPLICANT:  WASCO, WILMA
;     TITLE OF INVENTION:  Genetic Alterations Related To Familial
;     TITLE OF INVENTION:  Alzheimer's Disease
;     NUMBER OF SEQUENCES:  42
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
;       STREET:  1100 NEW YORK AVENUE, SUITE 600
;       CITY:  WASHINGTON
;       STATE:  DC
;       COUNTRY:  USA
;       ZIP:  20005-3934
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/706,344C
;       FILING DATE:  30-AUG-1996
;       CLASSIFICATION:  536
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  60/003,054
;       FILING DATE:  31-AUG-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  KIM, JUDITH U.
;       REGISTRATION NUMBER:  40,679
;       REFERENCE/DOCKET NUMBER:  0609.41B0001
```

```
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  202-371-2600
;       TELEFAX:  202-371-2540
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2765 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  249..1649
US-08-706-344C-3


alignment_scores:
              Quality:    6.00              Length:       6
                Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_215_222 x US-08-706-344C-3    ..

 Align seq 1/1  to: US-08-706-344C-3  from: 1  to: 2765

        1 MetLysGluArgThrSer 6
          ||||||||||||||||||
      185 ATGAAAGAAAGAACCTCA 202

seq_name: /cgn2_6/ptodata/1/ina/US_08_706_344C_3.seq; M_type: 3; Tot_seq: 27;

seq_documentation_block:
; Sequence 27, Application US/08706344C
; Patent No. 6248555
;  GENERAL INFORMATION:
;     APPLICANT:  TANZI, RUDOLPH
;     APPLICANT:  WASCO, WILMA
;     TITLE OF INVENTION:  Genetic Alterations Related to Familial
;     TITLE OF INVENTION:  Alzheimer's Disease
;     NUMBER OF SEQUENCES:  42
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
;       STREET:  1100 NEW YORK AVENUE, SUITE 600
;       CITY:  WASHINGTON
;       STATE:  DC
;       COUNTRY:  USA
;       ZIP:  20005-3934
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/706,344C
;       FILING DATE:  30-AUG-1996
;       CLASSIFICATION:  536
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  60/003,054
;       FILING DATE:  31-AUG-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  KIM, JUDITH U.
;       REGISTRATION NUMBER:  40,679
;       REFERENCE/DOCKET NUMBER:  0609.4180001
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  202-371-2600
;       TELEFAX:  202-371-2540
;   INFORMATION FOR SEQ ID NO:  27:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2765 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  249..1649
US-08-706-344C-27


alignment_scores:
              Quality:    6.00              Length:       6
                Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_215_222 x US-08-706-344C-27    ..

 Align seq 1/1  to: US-08-706-344C-27  from: 1  to: 2765

        1 MetLysGluArgThrSer 6
          ||||||||||||||||||
      185 ATGAAAGAAAGAACCTCA 202

seq_name: /cgn2_6/ptodata/1/ina/US_08_706_344C_27.seq; M_type: 3; Tot_seq: 29;

seq_documentation_block:
; Sequence 29, Application US/08706344C
; Patent No. 6248555
;  GENERAL INFORMATION:
;     APPLICANT:  TANZI, RUDOLPH
;     APPLICANT:  WASCO, WILMA
;     TITLE OF INVENTION:  Genetic Alterations Related to Familial
;     TITLE OF INVENTION:  Alzheimer's Disease
;     NUMBER OF SEQUENCES:  42
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
;       STREET:  1100 NEW YORK AVENUE, SUITE 600
;       CITY:  WASHINGTON
;       STATE:  DC
;       COUNTRY:  USA
;       ZIP:  20005-3934
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/706,344C
;       FILING DATE:  30-AUG-1996
;       CLASSIFICATION:  536
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  60/003,054
;       FILING DATE:  31-AUG-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  KIM, JUDITH U.
;       REGISTRATION NUMBER:  40,679
;       REFERENCE/DOCKET NUMBER:  0609.4180001
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  202 371 2600
;       TELEFAX:  202-371-2540
;   INFORMATION FOR SEQ ID NO:  29:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2765 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  249..1649
US-08-706-344C-29


alignment_scores:
              Quality:    6.00              Length:       6
```

```
              Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-126-945B-2_COPY_215_222 x US-08-706-844C-29    ..

 Align seq 1/1  to: US-08-706-844C-29  from: 1  to: 2765

        1 MetLysCluArgThrSer 6
          ||||||||||||||||||
      185 ATGAAACAAAGAACCTCA 202

Seq name: /projects/pstalign/CGI_bin/CGMB_seq?...us-09-126-945b-2.seq:US-08-706-844C-29

Seq documentation block:
; Sequence 41, Application US/08706844C
; Patent No. 6248555
;  GENERAL INFORMATION:
;    APPLICANT:  TANZI, RUDOLPH
;    APPLICANT:  WASCO, WILMA
;    TITLE OF INVENTION:  Genetic Alterations Related To Familial
;    TITLE OF INVENTION:  Alzheimer's Disease
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/706,844C
;      FILING DATE:  30-AUG-1996
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/003,054
;      FILING DATE:  31-AUG-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KIM, JUDITH U.
;      REGISTRATION NUMBER:  40,679
;      REFERENCE/DOCKET NUMBER:  0609.4180001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  41:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2765 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  249..1649
US-08-706-844C-41


Alignment scores:
             Quality:    6.00              Length:       6
               Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-126-945B-2_COPY_215_222 x US-08-706-844C-41    ..

 Align seq 1/1  to: US-08-706-844C-41  from: 1  to: 2765

        1 MetLysCluArgThrSer 6
```

```
          ||||||||||||||||||
      185 ATGAAACAAAGAACCTCA 202

Seq name: /projects/pstalign/CGI_bin/CGMB_seq?...us-09-126-945b-2.seq:US-08-967-101-1

Seq documentation block:
; Sequence 1, Application US/08967101
; Patent No. 5840540
;  GENERAL INFORMATION:
;    APPLICANT:  ST. GEORGE-HYSLOP, PETER H
;    APPLICANT:  ROMMENS, JOHANNA M
;    APPLICANT:  FRASER, PAUL E
;    TITLE OF INVENTION:  GENETIC SEQUENCES AND PROTEINS RELATED
;    TITLE OF INVENTION:  TO ALZHEIMER'S DISEASE
;    NUMBER OF SEQUENCES:  183
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  High Street Tower - 125 High Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/967,101
;      FILING DATE:  18-NOV-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/592,541
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fitcher, Edmund R.
;      REGISTRATION NUMBER:  30,959
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7000
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2791 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-967-101-1


Alignment scores:
             Quality:    6.00              Length:       6
               Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-126-945B-2_COPY_215_222 x US-08-967-101-1    ..

 Align seq 1/1  to: US-08-967-101-1  from: 1  to: 2791

        1 MetLysCluArgThrSer 6
          |||| |||||||:||||||
      185 ATGAAAGAAAGAACCTCA 203

Seq name: /projects/pstalign/CGI_bin/CGMB_seq?...us-09-126-945b-2.seq:US-08-967-101-133

Seq documentation block:
; Sequence 133, Application US/08967101
; Patent No. 5840540
;  GENERAL INFORMATION:
;    APPLICANT:  ST. GEORGE-HYSLOP, PETER H
;    APPLICANT:  ROMMENS, JOHANNA M
;    APPLICANT:  FRASER, PAUL E
;    TITLE OF INVENTION:  GENETIC SEQUENCES AND PROTEINS RELATED
```

```
;     TITLE OF INVENTION:  TO ALZHEIMER'S DISEASE
;     NUMBER OF SEQUENCES:  183
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;       STREET:  High Street Tower  125 High Street
;       CITY:  Boston
;       STATE:  Massachusetts
;       COUNTRY:  U.S.A.
;       ZIP:  02110
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/967,101
;       FILING DATE:  10-NOV-1997
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/592,541
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Pitcher, Edmund R.
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 248-7000
;       TELEFAX:  (617) 248-7100
;   INFORMATION FOR SEQ ID NO:  133:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2791 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-967-101-133


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_215_222 x US-08-967-101-133    .

 Align seq 1/1  to: US-08-967-101-133  from: 1  to: 2791

       1 MetLysGluArgThrSer 6
         ||||||||||||||||||
     185 ATGAAAGAAAGAACTCA 202

Seq name: /tmp2_fpt_data/2_fma_5B_COMB_seq-US-08-592-541-1

seq_documentation_block:
; Sequence 1, Application US/08592541
; Patent No. 5986054
;  GENERAL INFORMATION:
;     APPLICANT:  ST. GEORGE-HYSLOP, PETER H
;     APPLICANT:  ROMMENS, JOHANNA M
;     APPLICANT:  FRASER, PAUL E
;     TITLE OF INVENTION:  GENETIC SEQUENCES AND PROTEINS RELATED
;     TITLE OF INVENTION:  TO ALZHEIMER'S DISEASE
;     NUMBER OF SEQUENCES:  183
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;       STREET:  High Street Tower - 125 High Street
;       CITY:  Boston
;       STATE:  Massachusetts
;       COUNTRY:  U.S.A.
;       ZIP:  02110
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/592,541
;       FILING DATE:
;       CLASSIFICATION:  800
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Pitcher, Edmund R.
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 248-7000
;       TELEFAX:  (617) 248-7100
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2791 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-592-541-1


alignment_scores:
            Quality:    6.00              Length:       6
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_215_222 x US-08-592-541-1    .

 Align seq 1/1  to: US-08-592-541-1  from: 1  to: 2791

       1 MetLysGluArgThrSer 6
         ||||||||||||||||||
     186 ATGAAAGAAAGAACTCA 203

seq_name: /tmp2_fpt_data/2_fma_5B_COMB_seq-US-08-592-541-1

seq_documentation_block:
; Sequence 133, Application US/08592541
; Patent No. 5986054
;  GENERAL INFORMATION:
;     APPLICANT:  ST. GEORGE-HYSLOP, PETER H
;     APPLICANT:  ROMMENS, JOHANNA M
;     APPLICANT:  FRASER, PAUL E
;     TITLE OF INVENTION:  GENETIC SEQUENCES AND PROTEINS RELATED
;     TITLE OF INVENTION:  TO ALZHEIMER'S DISEASE
;     NUMBER OF SEQUENCES:  183
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;       STREET:  High Street Tower - 125 High Street
;       CITY:  Boston
;       STATE:  Massachusetts
;       COUNTRY:  U.S.A.
;       ZIP:  02110
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/592,541
;       FILING DATE:
;       CLASSIFICATION:  800
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Pitcher, Edmund R.
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 248-7000
;       TELEFAX:  (617) 248-7100
;   INFORMATION FOR SEQ ID NO:  133:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2791 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
```

```
;     MOLECULE TYPE:  cDNA
US 08 592-541 133


alignment_scores:
             Quality:     6.00              Length:       6
               Ratio:  1.000                Gaps:         0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US 09 126 945B 2_COPY_215_222 x US 08-592-541 133    ..

 Align seq 1/1  to: US-08-592-541 133  from: 1  to: 2791

        1 MetLysGluArgThrSer 6
          ||||||||||||||||||
      185 ATGAAAGAAAGAACCTCA 202
```

```
OM of: US-09-126-945B-2_COPY_215_222 to: EST:*  Out_format : pto

Date: Sep 14, 2001  7:47 PM

About: Results were produced by the GenCore software, version 4.5.
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL frame=_p2n.model -DEV xlp
 -Q="/cgn2_2/ftptmp/sp_n_1356+1764+45_3_out_21_0_1001_14651_14756/qil_poly_19611_1.fa"
 -DB=EST -QFMT=fastap -SUFFIX olip2n.rst -GAP=P 4.500
 -GAPEXT 0.050 -MINMATCH 0.100 -LOOPML 0.000 -LOOPEXT 0.000
 -QGAP=P 4.500 -QGAPEXT=0.050 -XGAP=P=60.000 -XGAPEXT 60.000
  YGAP=P 6.000 -YGAPEXT=7.000 -YGAP=P 60.000 -YGAPEXT 60.000
 -DELOP 6.000 -DELEXT=7.000 -START 1 -MATRIX oligo
 -TRANS human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE quality
 -THR_MIN 1  ALIGN 15  MODE LOCAL  OUTFMT pto  NORM OXT  MINLEN 3
 -MAXLEN 2000000000  USER US0912695P_CGNI_1_4645  NCPU 6
 -ICPU 3 -LONGLOG -NO_XLPXY -WAIT -THREADS 1

Search information block:
Query: US-09-126-945B-2_COPY_215_222
Query length: 8
Database: EST:*
Database sequences: 10228115
Database length: 431459454
Search time (sec): 9087.630000

WARN: XGAP=P and YGAP=P must be equal. Assuming YGAP=P XGAP=P 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000


Score_list:
Sequence                Strd Orig    ZScore    EScore Len  ! Documentation ..
qb_qss27:AZ457779         +    8.00  151.32     26.10    449  ! AZ457779 IM0261118F Mouse 10kb
qb_est78:BE757615         +    8.00  150.14     30.37    539  ! BE757615 212104 MAGC 2B0V Bos t
qb_est96:BG085247         +    8.00  149.54     32.78    591  ! BG085247 H41b0812 5 NIA Mouse 1
qb_est97:BG179568         +    8.00  146.20     59.82    991  ! BG179568 60232s182F1 NIH_MGC_>
qb_est89:BF579116         +    8.00  141.46     92.46   2064  ! BF579116 60209605OF1 NCI_CGAP_
qb_est105:H53378          -    7.00  136.42    176.45    261  ! H53378 yr86b04.r1 Soares_pineal
qb_est103:C42310          -    7.00  135.52    192.93    300  ! C42310 C4231v Yuji Kohara unpub
qb_est103:C38692          +    7.00  135.52    192.93    300  ! C38692 C38692_ Yuji Kohara unpub
qb_est4:AA212628          -    7.00  134.04    249.21    377  ! AA212628 ms14h10.r1 Soares_thym
qb_est4:AA229911          +    7.00  133.99    249.79    380  ! AA229911 ms37b08.s1 NCI_CGAP_Fe
qb_est49:AW583416         +    7.00  133.73    249.17    396  ! AW583416 ia02408.x1 Barstead Colo
qb_est25:A1821500         +    7.00  133.25    264.72    426  ! A1821500 ms37b08.x1 NCI_CGAP_Fe
qb_est96:BG097521         -    7.00  132.74    282.64    461  ! BG097521 PST461456 potato tuber
qb_est25:A1761640         -    7.00  132.70    284.15    464  ! A1761640 ms37b08.y1 NCI_CGAP_Fe
qb_est70:BE073809         -    7.00  132.26    299.81    497  ! BE073809 RC0-BI0561-240200-034
qb_est71:A1782049         -    7.00  132.22    302.72    500  ! A1782049 EST2062928 tomato susce
qb_est53:AW034223         -    7.00  131.92    318.27    532  ! AW034223 EST2060956 tomato fruit
qb_est70:BE073804         -    7.00  131.38    336.52    569  ! BE073804 RC0-BI0561-240200-034
qb_qss34:AZ883069         +    7.00  131.37    337.01    570  ! AZ883069 RPCI-24-204K2.1V RPCI
qb_qss29:AZ015410         +    7.00  130.99    351.09    605  ! AZ015410 RPCI-24-476C3.2T RPCI
qb_est102:BG593781        -    7.00  130.87    356.41    616  ! BG593781 EST1492459 cSTS Solanum
qb_est82:BF051317         -    7.00  130.84    360.86    619  ! BF051317 EST1466492 tomato shoot
qb_est97:BG127861         -    7.00  130.80    362.79    623  ! BG127861 EST1473507 tomato shoot
qb_qss33:AZ886253         -    7.00  130.79    363.28    624  ! AZ886253 JM0141B34F Mouse 10kb
qb_est74:BE436155         -    7.00  130.58    372.91    644  ! BE436155 EST1407233 tomato break
qb_est41:AV760922         +    7.00  130.55    374.34    647  ! AV760922 AV760922 MDS Homo sapi
qb_est102:BG590033        -    7.00  130.10    396.76    694  ! BG590033 PST497875 P. hybostans
qb_est95:BG018798         -    7.00  129.80    412.34    727  ! BG018798 dag56h03.y1 Weizmann
qb_est102:BG600695        -    7.00  129.78    413.28    729  ! BG600695 EST505602 cSTS Solanum
qb_est102:BG598545        -    7.00  129.77    413.75    730  ! BG598545 PST503445 cSTS S.lanum
qb_qss28:AZ557293         +    7.00  129.51    427.80    760  ! AZ557293 RPCI-23-214P20.1T RPCI
qb_est96:BG072422         -    7.00  129.41    438.97    784  ! BG072422 H41b0812 3 NIA Mouse 1
qb_qss22:AZ183875         +    7.00  129.01    456.51    822  ! AZ183875 SP_1002_A1_E08_SP6 Str
qb_qss4:CNS028LR          +    7.00  128.83    466.65    844  ! AL186120 Tetraodon nigroviridis
qb_qss4:CNS019JN          +    7.00  128.12    511.60    943  ! AL245300 Tetraodon nigroviridis
qb_est84:BF124921         +    7.00  127.95    522.82    968  ! BF124921 601752728F1 NIH_MGC_2
qb_qss4:CNS02K3H          -    7.00  127.10    582.59   1103  ! AL201014 Tetraodon nigroviridi
qb_est70:BE140640         -    6.00  128.66    905.70    109  ! BE140640 RC0-HT0014-310599-021
qb_est104:D76992          -    6.00  128.21    969.18    117  ! D76992 M0393008 mouse embryonal
qb_qss23:AZ268931         -    6.00  122.88   1.0e+03    123  ! AZ268931 RPCI-23-444D23.T1 R
qb_est15:A1620552         -    6.00  122.53   1.0e+03    130  ! A1620552 ub24b04.r1 Soares t
qb_est77:BE665137         -    6.00  122.53   1.0e+03    130  ! BE665137 15d5r5 MAGC 4B0V Bo
qb_est13:AA929594         -    6.00  122.19   1.1e+03    137  ! AA929594 zz56c12.r1 Soares m
qb_est2:AA090727          -    6.00  121.56   1.2e+03    151  ! AA090727 yy0143.seq.F Human
qb_est95:BF998853         -    6.00  121.51   1.2e+03    152  ! BF998853 MK3 cNo152-Tollo c

seq_name: qb_qss27:AZ457779

seq_documentation_block:
LOCUS       AZ457779          449 bp    DNA              GSS       04-OCT-2000
DEFINITION  IM0261118F Mouse 10kb plasmid UGGCIM library Mus musculus genomic
            clone UGGCIM0261118 F, DNA sequence.
ACCESSION   AZ457779
VERSION     AZ457779.1  GI:10615904
KEYWORDS    GSS.
SOURCE      mouse mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,R., Londadeo,S., Mahmoud,M., Meenon,E., Pederson,J., Reilly,
            M., Rose,M., Rose,R., St Rose,S., Tinney,A., Y.-H. Nie,Brian,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000  Std Error: 0.00
            Plate: 0261  row: I  column: 18
            Seq primer: CGTTGIAAAACGACGGCCAGI
            Class: plasmid ends
            High quality sequence stop: 449.
FEATURES             Location/Qualifiers
     source          1..449
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UGGCIM0261118"
                     /clone_lib="Mouse 10kb plasmid UGGCIM library"
                     /sex="Male"
                     /lab_host="E. coli strain XL10-Gold, 11 resistant, F'"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained from the Jackson
                     Laboratory Mouse DNA Resource
                     [URL: http://www.jax.org/resources/documents/cmr/]. DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       76 a    110 c    140 g    129 t
ORIGIN
```

```
Alignment scores:
            Quality:     8.00              Length:       8
              Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

Alignment block:
 US-09-126-945B-2-COPY-215_222 x AZ457729  ..

 Align seq 1/1  to: AZ457729  from: 1  to: 449

        1 MetLysGluArgThrSerProGly 8
          ||||||||||||||||||||||||
      412 ATGAAGGAGAGGACCTCCCCTGGG 435

Seq name: qb_est78:BE757615

Seq documentation block:
LOCUS        BE757615       539 bp    mRNA             EST       15-SEP-2000
DEFINITION   Z12104 MARC 2B0V Bos taurus cDNA 5', mRNA sequence.
ACCESSION    BE757615
VERSION      BE757615.1  GI:10171607
KEYWORDS     EST.
SOURCE       cow.
  ORGANISM   Bos taurus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
             Bovidae; Bovinae; Bos.
REFERENCE    1  (bases 1 to 539)
  AUTHORS    Smith,T.P.L., Casas,E., Stone,R.T., Beaton,M.P., Grosse,W.M.,
             Bennett,G.A., Fahrenkrug,S.C., Ensley,B.A., Kappes,S.A., Lesperid
             ,W.W. and Keele,J.W.
  TITLE      Design and use of four pooled tissue normalized cDNA libraries for
             EST discovery in cattle
  JOURNAL    Unpublished (2000)
COMMENT      Contact: Smith TPL
             USDA, ARS, US Meat Animal Research Center
             PO Box 166, Clay Center, NE 68933-0166, USA
             Tel: 402-762-4466
             Fax: 402-762-4490
             Email: smith@email.marc.usda.gov
             Single pass sequencing. Bases called and alt trimmed with phred
             v0.980904.e. Vector identified by cross_match with the minscore 19
             and minmatch 12 options.
             PCR Primers
             FORWARD: AGGAAACAGCTATGACCAT
             BACKWARD: GTTTTCCCAGTCACGACG
             Plate: 64  row: D  column: 13
             Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1..539
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 2B0V"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from testis, thymus,
                     semitendonosus muscle, longissimus muscle, pancreas,
                     adrenal, and endometrium."
BASE COUNT       98 a    168 c    194 g    80 t
ORIGIN


Alignment scores:
            Quality:     8.00              Length:       8
              Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

Alignment block:
 US-09-126-945B-2-COPY-215_222 x BE757615

 Align seq 1/1  to: BE757615  from: 1  to: 539

        1 MetLysGluArgThrSerProGly 8
          ||||||||||||||||||||||||
      495 ATGAAGGAGAGGACCTCCCCTGGG 418

Seq name: qb_est78:BG085247

Seq documentation block:
LOCUS        BG085247       591 bp    mRNA             EST       26-JAN-2001
DEFINITION   H31D8C12 5', NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
             H31D8C12 5', mRNA sequence.
ACCESSION    BG085247
VERSION      BG085247.1  GI:12567811
KEYWORDS     EST.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 591)
  AUTHORS    Karpui,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Jaradat,S.A., Tanaka
             ...S., Carter,M.G. and Ko,M.S.H.
  TITLE      Verification and initial annotation of NIA mouse 15K cDNA clone set
  JOURNAL    Unpublished (2001)
COMMENT      Other ESTs: H31D8C12 3'
             Contact: George J. Karpui
             Laboratory of Genetics
             National Institute on Aging/National Institutes of Health
             333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
             Email: cdnalqsun.grc.nia.nih.gov
             This clone set has been freely distributed to the community. Please
             visit http://lgsun.grc.nia.nih.gov/cDNA/15k.html for details.
             Plate: H31D  row: C  column: 12
             Seq primer:  21M13 Reverse
             High quality sequence stop: 591
             PolyA No.
FEATURES             Location/Qualifiers
     source          1..591
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="H31D8C12"
                     /clone_lib="NIA Mouse 15K cDNA Clone Set"
                     /sex="Mixed from a variety of cDNA libraries"
                     /dev_stage="Clones arrayed from a variety of cDNA
                     libraries"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
                     clone is among a rearrayed set of 15,247 clones from 11
                     embryo cDNA libraries (including preimplantation stage
                     embryos from unfertilized egg to blastocyst, embryonic
                     part of E7.5 embryos, extraembryonic part of E7.5 embryos
                     and E12.5 female mesonephros/gonad) and one newborn
                     ovary cDNA library. Average insert size 1.5 Kb. All
                     source libraries are cloned unidirectionally with oligo(dT
                     ) Not primers. References include: (1) Genome-wide
                     expression profiling of mid-gestation placenta and embryo
                     using a 15,000 mouse developmental cDNA microarray, 2000,
                     Proc. Natl. Acad. Sci. U.S.A. 97: 9127-9132; (2)
                     Large-scale cDNA analysis reveals phased gene-expression
                     patterns during preimplantation mouse development, 2000,
                     Development, 127: 1737-1749; (3) Genome-wide mapping of
                     uncloned transcripts from extraembryonic tissue of
                     7.5-day mouse embryos reveals enrichment in the t-complex
                     and under-representation on the X chromosome, 1998, Hum
                     Mol Genet 7: 1967-1978."
BASE COUNT      135 a    168 c    175 g    113 t
ORIGIN


Alignment scores:
            Quality:     8.00              Length:       8
              Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000
```

```
alignment_block:
 US-09-126-945B-2_COPY_215_222 x BG085247     .

 Align seq 1/1  to: BG085247  from: 1  to: 247

        1 MetLysGluArgThrSerProGly 8
          ||||||||||||||||||||||||
      223 ATGAAGGAGAGGACCTCGCCTGGG 246

seq_name: gb_est97:BG179568

seq_documentation_block:
LOCUS        BG179568       991 bp    mRNA            EST       06 FEB 2001
DEFINITION  602928182F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4429413 5',
            mRNA sequence.
ACCESSION   BG179568
VERSION     BG179568.1  GI:12686271
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 991)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DIP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10181  row: d  column: 22
            High quality sequence stop: 655.
FEATURES             Location/Qualifiers
     source          1. .991
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4429413"
                     /clone_lib="NIH_MGC_91"
                     /tissue_type "adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: prostate; Vector: pCMV SPORT6; Site_1: Not I;
                     Site_2: Sal I; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.4 kB. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      190 a    351 c    305 g    145 t
ORIGIN


alignment_scores:
            Quality:     8.00             Length:       8
              Ratio:    1.000              Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_215_222 x BG179568     .

 Align seq 1/1  to: BG179568  from: 1  to: 991

        1 MetLysGluArgThrSerProGly 8
          ||||||||||||||||||||||||
      546 ATGAAAGAGCGGACTTCACCTGGG 569

seq_name: gb_est89:BF579116

seq_documentation_block:
LOCUS        BF579116      2064 bp    mRNA            EST       12 DEC 2000
DEFINITION  602896050F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4215901 5',
            mRNA sequence
ACCESSION   BF579116
VERSION     BF579116.1  GI:11652828
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2064)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9792  row: d  column: 14
            High quality sequence stop: 645.
FEATURES             Location/Qualifiers
     source          1. .2064
                     /organism "Mus musculus"
                     /strain "FVB/N"
                     /db_xref "taxon:10090"
                     /clone "IMAGE:4215901"
                     /clone_lib "NCI_CGAP_Co24"
                     /lab_host "DH10B (T1 phage resistant)"
                     /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: Not I;
                     Site_2: Sal I; Cloned unidirectionally.  Primer: oligo dT.
                     Average insert size 1.6 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      624 a    621 c    620 g    199 t
ORIGIN


alignment_scores:
            Quality:     8.00             Length:       8
              Ratio:    1.000              Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_215_222 x BF579116     .

 Align seq 1/1  to: BF579116  from: 1  to: 2064

        1 MetLysGluArgThrSerProGly 8
          ||||||||||||||||||||||||
      595 ATGAAGGAGAGGACCTCGCCTGGG 618

seq_name: gb_est105:H53478

seq_documentation_block:
LOCUS        H53478         261 bp    mRNA            EST       20 SEP 1995
DEFINITION  yy86d04.r1 Soares pineal_gland_N3HPG Homo sapiens cDNA clone
            IMAGE:254151 5' similar to gb:L02932 PEROXISOME PROLIFERATOR
            ACTIVATED RECEPTOR ALPHA (HUMAN);, mRNA sequence.
ACCESSION   H53478
VERSION     H53478.1  GI:993525
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 261)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
            ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
            ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
            Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Wiremaskis,E.,
```

```
            Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807 828 (1996)
  MEDLINE   97044478
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, Mo 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 696
            High quality sequence stops: 54
            Source: IMAGE Consortium, LLNL
            This clone is available royalty free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 696  Std Error: 0.00
            Seq primer: M13RPI
            High quality sequence stop: 54.
FEATURES             Location/Qualifiers
     source          1..261
                     /organism="Homo sapiens"
                     /db_xref="GDB:086114"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:241151"
                     /clone_lib="Soares pineal gland N3HPG"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: pineal gland; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site 1: Not I; Site 2: Eco RI;
                     1st strand cDNA was primed with a Not I oligo(dT) primer
                     [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTT 3']
                     , double-stranded cDNA was size selected, ligated to Eco
                     RI adapters (Pharmacia), digested with Not I and cloned
                     into the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library constructed by Bento Soares and
                     M.Fatima Bonaldo. "
BASE COUNT        75 a      54 c      52 g      81 t
ORIGIN


alignment scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000               Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2_COPY_215_222 x H53478/rev   ..

 Align seq 1/1  to reverse of: H53478  from: 1  to: 261

        1 MetLysGluArgThrSerPro 7
          |||||||||||||||||||||
      244 ATGAAGGAAAGAACAAGCCCT 224

seq name: qb-est104:C32410

seq documentation block:
LOCUS       C32410          300 bp    mRNA             EST       18-OCT-1999
DEFINITION  C32410 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite embryo
            Caenorhabditis elegans cDNA clone yk327q1 3', mRNA sequence.
ACCESSION   C32410
VERSION     C32410.1  GI:2364106
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
            ; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 300)
  AUTHORS   Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A., Sano
            ,M., Miyata,A. and Nishiqaki,A.
  TITLE     Expression map of the C.elegans genome
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81 559-81 6854
            Fax: 81 559-81 6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1..300
                     /organism="Caenorhabditis elegans"
                     /strain="N2"
                     /db_xref="taxon:6239"
                     /clone="yk327q1"
                     /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
                     hermaphrodite embryo"
                     /sex="hermaphrodite"
                     /dev_stage="embryo"
BASE COUNT       105 a      39 c      71 g      78 t       7 others
ORIGIN


alignment_scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000               Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2_COPY_215_222 x C32410   ..

 Align seq 1/1  to: C32410  from: 1  to: 300

        1 MetLysGluArgThrSerPro 7
          |||||||||||||||||||||
      105 ATGAAAGAGAGAACNICICCA 125

seq name: qb-est104:C38602

seq documentation block:
LOCUS       C38602          300 bp    mRNA             EST       18-OCT-1999
DEFINITION  C38602 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite embryo
            Caenorhabditis elegans cDNA clone yk495l4 3', mRNA sequence.
ACCESSION   C38602
VERSION     C38602.1  GI:2374839
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
            ; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 300)
  AUTHORS   Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A., Sano
            ,M., Miyata,A. and Nishiqaki,A.
  TITLE     Expression map of the C.elegans genome
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81 559 81 6854
            Fax: 81 559-81 6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1..300
                     /organism="Caenorhabditis elegans"
                     /strain="N2"
                     /db_xref="taxon:6239"
                     /clone="yk495l4"
                     /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
                     hermaphrodite embryo"
                     /sex="hermaphrodite"
                     /dev_stage="embryo"
BASE COUNT       110 a      38 c      75 g      76 t       1 others
ORIGIN


alignment scores:
```

```
            Quality:     7.00              Length:       7
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_215_222 x C38602    .

 Align seq 1/1  to: C38602  from: 1  to: 300

        1 MetLysGluArgThrSerPro 7
          |||||||||||||||||||
      161 ATGAAAGAGACAATCTCCA 121

Seq_name: gb_est3:AA212628

seq_documentation_block:
LOCUS        AA212628       377 bp    mRNA            EST       19-FEB-1997
DEFINITION   mm44b10.r1 Soares_thymus_2NbMI Mus musculus cDNA clone IMAGE:642307
             5', mRNA sequence.
ACCESSION    AA212628
VERSION      AA212628.1  GI:1811273
KEYWORDS     EST.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 377)
  AUTHORS    Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Iobbagy,I.,
             Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
             Schellenberg,K., Stoptoe,M., Tan,F., Underwood,K., Moore,B.,
             Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
             Waterston,R.
  TITLE      The WashU-HHMI Mouse EST Project
  JOURNAL    Unpublished (1996)
COMMENT      Contact: Marra M/Mouse EST Project
             WashU-HHMI Mouse EST Project
             Washington University School of Medicine
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: mouseest@watson.wustl.edu
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             MGI:394299
             Seq primer: -28ml3 rev2 ET from Amersham
             High quality sequence stop: 351.
FEATURES             Location/Qualifiers
     source          1. .377
                     /organism="Mus musculus"
                     /strain="SPRET/Ei"
                     /db_xref="taxon:10090"
                     /clone "IMAGE:642307"
                     /clone_lib "Soares_thymus_2NbMI"
                     /sex="male"
                     /tissue_type="Thymus"
                     /dev_stage "4 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. RNA
                     provided by Dr. Bertrand Jordan. Library went through two
                     rounds of normalization, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT       116 a      80 c      82 g      99 t
ORIGIN


alignment_scores:
            Quality:     7.00              Length:       7
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_215_222 x AA212628_rev    .

 Align seq 1/1  to reverse of: AA212628  from: 1  to: 377

        2 LysGluArgThrSerProGly 8
          |||||||||||||||||||
      271 AAAGAAAGGACAATCCCGGGC 251

Seq_name: gb_est4:AA229911

seq_documentation_block:
LOCUS        AA229911       380 bp    mRNA            EST       21-AUG-1997
DEFINITION   ie47b08.s1 NCI_CGAP_Pr2 Homo sapiens cDNA clone IMAGE:1010367, mRNA
             sequence.
ACCESSION    AA229911
VERSION      AA229911.1  GI:1852206
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 380)
  AUTHORS    [illegible]
  TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
  JOURNAL    Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui, M.D.
             , Michael Emmert Buck, M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             [illegible]
             Insert Length: 337  Std Error: 0.00
             Seq primer: -41ml3 fwd. ET from Amersham
             High quality sequence stop: 357.
FEATURES             Location/Qualifiers
     source          1. .380
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone "IMAGE:1010367"
                     /clone_lib "NCI_CGAP_Pr2"
                     /sex "Male"
                     /dev_stage "45 years old"
                     /lab_host "DH10B"
                     /note "Vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st
                     strand cDNA was primed with oligo(dT)17 on 50 ng of
                     DNAse treated, total cellular RNA obtained from 5,000 to
                     [illegible] microdissected preneoplastic cells
                     histologically-determined to be prostatic intraepithelial
                     neoplasia 2 (PIN2) cells. Double stranded cDNA was
                     ligated to EcoRI adaptors, 5 cycles of PCR applied to the
                     cDNA with an adaptor-specific primer, and the resulting
                     PCR product subcloned into pAMP10 by the UDG cloning
                     method (Life Technologies). Average insert size is 400
                     bp. NOTE: Not directionally cloned. This library was
                     constructed by David Krizman."
BASE COUNT       105 a      78 c      73 g     124 t
ORIGIN


alignment_scores:
            Quality:     7.00              Length:       7
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000
```

```
alignment block:
 US-09-126-945B-2-COPY_215_222 x AA229911  ..

 Align seq 1/1  to: AA229911  from: 1  to: 380

        1 MetLysGluArgThrSerPro 7
          |||||||||||||||||||||
      305 ATGAAGGAAAGAACAAGCCCT 325

seq name: qb_est49:AW583416

seq documentation block:
LOCUS        AW583416        496 bp    mRNA             EST       17 MAR 2000
DEFINITION   ia02d08.x1 Human Pancreatic Islets Homo Sapiens cDNA 3' similar to
             qb:X17206 40S RIBOSOMAL PROTEIN S4 (HUMAN); qb:M20642 Mouse LLRep3
             protein mRNA from a repetitive element, complete (MOUSE);, mRNA
             sequence.
ACCESSION    AW583416
VERSION      AW583416.1  GI:7266771
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 496)
  AUTHORS    Melton,D., Meadows,A., Clifton,S., Hillier,L., Marra,M., Pape,S.,
             Wylie,T., Martin,J., Blistain,A., Schmitt,A., Theising,B., Ritter
             ,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M., McCann,R., Colo
             ,R., Tsagareishvili,R., Williams,L., Jackson,Y. and Bowers,Y.
  TITLE      Wash0 Harvard Pancreas EST Project
  JOURNAL    Unpublished (2000)
COMMENT      Contact: Douglas Melton
             Wash0 Harvard Pancreas EST Project
             Harvard University, Howard Hughes Medical Institute
             Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
             MA 02138
             Tel: 617 495-1812
             Fax: 617 495-8557
             Email: dmelton@biohp.harvard.edu, meadows@tas.harvard.edu
             Library was constructed by Douglas Melton, Ph.D. DNA sequencing by:
             Washington University Genome Sequencing Center For information on
             obtaining a clone please contact: Adimika Meadows
             (meadows@tas.harvard.edu)
             Seq primer:  40UP from Gibco
             High quality sequence stop: 60.
FEATURES             Location/Qualifiers
     source          1. .496
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone_lib "Human Pancreatic Islets"
                     /tissue_type "Islets of Langerhans"
                     /dev_stage "Adult"
                     /lab_host "DH10B"
                     /note "Organ: Pancreas; Vector: pSPORT1; Site 1: Not I;
                     Site_2: Sal I; Library constructed using SuperScript
                     Plasmid Library kit (Life Technologies). cDNA made by
                     oligo dT priming.  Size selected by column fractionation;
                     average insert size 1.08 kb. Primary library,
                     unamplified."
BASE COUNT       88 a    118 c    108 g     82 t
ORIGIN


alignment scores:
            Quality:    7.00            Length:       7
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_215_222 x AW583416  ..

 Align seq 1/1  to: AW583416  from: 1  to: 496

        2 LysGluArgThrSerProGly 8
          |||| ||||||||||||||||
      421 AAAGAAAGAACCTCTCCAGGG 441

seq name: qb_est25:AI821500

seq documentation block:
LOCUS        AI821500        426 bp    mRNA             EST       13 DEC 1999
DEFINITION   ne87h08.y5 NCI_CGAP_Pr2 Homo Sapiens cDNA clone IMAGE:1010367, mRNA
             sequence.
ACCESSION    AI821500
VERSION      AI821500.1  GI:5440579
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 426)
  AUTHORS    NCI-CGAP Project (http://www.ncbi.nlm.nih.gov/ncicgap).
  TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
  JOURNAL    Unpublished (1997)
COMMENT      Other_ESTs: ne87h08.y5
             Contact: Robert Strausberg, Ph.D.
             Email: cgaps@mail.nih.gov
             Tissue Procurement: W. Marston Linehan, M.D., Kohtaro Ohnuki, M.D.
             , Michael Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Genome Systems Inc., Creg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             www.bio.llnl.gov/bbrp/image/image.html

             This read is a RESEQUENCE of a previously sequenced human clone
             Original clone citation: Wash0 NCI EST Project
             This read has been verified (found to hit its original self in the
             correct orientation)
             Insert Length: 437  Std Error: 0.00
             Seq primer:  40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .426
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone "IMAGE:1010367"
                     /clone_lib "NCI_CGAP_Pr2"
                     /sex "Male"
                     /dev_stage "45 years old"
                     /lab_host "DH10B"
                     /note "Vector: pAMP10; Site 1: NotI; Site 2: EcoRI; 1st
                     strand cDNA was primed with oligo(dT)17 on 50 ng of
                     DNAse treated, total cellular RNA obtained from 5,000 to
                     ,000 microdissected preneoplastic cells
                     histologically determined to be prostatic intraepithelial
                     neoplasia 2 (PIN2) cells.  Double stranded cDNA was
                     ligated to EcoRI adaptors, 5 cycles of PCR applied to the
                     cDNA with an adaptor-specific primer, and the resulting
                     PCR product subcloned into pAMP10 by the UDG cloning
                     method (Life Technologies).  Average insert size is 600
                     bp. NOTE: Not directionally cloned.  This library was
                     constructed by  David Krizman."
BASE COUNT      117 a     84 c     80 g    146 t
ORIGIN


alignment scores:
            Quality:    7.00            Length:       7
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY_215_222 x AI821500  ..
```

```
 Align seq 1/1  to: AI821500  from: 1  to: 426

       1 MetLysGlnArgThrSerPro 7
         |||||||||||||||||||||
     329 ATGAAGCAAAGAACAAGCCCT 349

Seq_name: qb_est96:BG097521

Seq_documentation_block:
LOCUS        BG097521       461 bp    mRNA            EST       29-JAN-2001
DEFINITION  EST461956 potato leaves and petioles Solanum tuberosum cDNA clone
            cSTB47121 5' sequence, mRNA sequence.
ACCESSION   BG097521
VERSION     BG097521.1  GI:12587472
KEYWORDS    EST
SOURCE      potato.
  ORGANISM  Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE   1  (bases 1 to 461)
  AUTHORS   van der Hoeyen,P.S., Borzerides,I., Bodis,I.E., Liang,F., Cho,J.,
            Utterback,T., Hansen,C.L., Doan,B., Bouqri,O., Buell,C.R., Ronning
            ,C.M., Fry,W.E., Tanksley,S.D. and Baker,B.
  TITLE     Generation of ESTs from potato leaves and petioles
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cathy Ronning
            The Institute for Genomic Research
            For clone info: please contact Research Genetics, Libraries
            Division tel 1-800-711-6195, email cdna@resgen.com.
FEATURES             Location/Qualifiers
     source          1..461
                     /organism="Solanum tuberosum"
                     /cultivar="Kennebec"
                     /db_xref="taxon:4113"
                     /clone="cSTB47121"
                     /clone_lib="potato leaves and petioles"
                     /tissue_type="leaflets and petioles"
                     /dev_stage="8 weeks old plant"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript SK( ); Site_1: EcoRI; Site_2:
                     XhoI; Tissue was supplied by Dr. Fry (Cornell University).
                     Leaflets and petioles were isolated from 8 week old
                     greenhouse grown plants. The plants were watered and
                     fertilized freely. The tissue was immediately frozen in
                     liquid nitrogen."
BASE COUNT      128 a    102 c     95 g    136 t
ORIGIN


alignment_scores:
            Quality:     7.00            Length:       7
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_215_222 x BG097521/rev   ..

 Align seq 1/1  to reverse of: BG097521  from: 1  to: 461

       1 MetLysGlnArgThrSerPro 7
         ||||||| |||||||||||||
     407 ATGAAACAGAGAACTAGCCCT 387

Seq_name: qb_est25:AI791640

Seq_documentation_block:
LOCUS        AI791640       464 bp    mRNA            EST       14-DEC-1999
DEFINITION  nc87h08.y5 NCI_CGAP_Pr2 Homo sapiens cDNA clone IMAGE:1010467, mRNA
            sequence.
ACCESSION   AI791640
VERSION     AI791640.1  GI:5339356
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 464)
  AUTHORS   [illegible]
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: nc87h08.s1
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui, M.D.
            , Michael Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: David B. Krizman, Ph.D.
            cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            [illegible]

            This read is a RESEQUENCE of a previously sequenced human clone
            original clone citation: WashU-NCI EST Project
            This read has been verified (found to hit its original self in the
            correct orientation)
            Insert length: 337  Std Error: 0.00
            Seq primer: -40RP from Gibco
            High quality sequence stop: 427.
FEATURES             Location/Qualifiers
     source          1..464
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1010467"
                     /clone_lib="NCI_CGAP_Pr2"
                     /sex="Male"
                     /dev_stage="45 years old"
                     /lab_host="DH10B"
                     /note="Vector: pAMP10; Site_1: NotI; Site_2: PvuII; 1st
                     strand cDNA was primed with oligo(dT)17 on 50 ng of
                     DNAse-treated, total cellular RNA obtained from 5,000 to
                     7,000 microdissected preneoplastic cells
                     histologically determined to be prostatic intraepithelial
                     neoplasia 2 (PIN2) cells. Double stranded cDNA was
                     ligated to EcoRI adaptors, 5 cycles of PCR applied to the
                     cDNA with an adaptor specific primer, and the resulting
                     PCR product subcloned into pAMP10 by the UDG cloning
                     method (Life Technologies). Average insert size is 600
                     bp. NOTE: Not directionally cloned. This library was
                     constructed by David Krizman."
BASE COUNT      133 a     98 c     89 g    144 t
ORIGIN


alignment_scores:
            Quality:     7.00            Length:       7
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_215_222 x AI791640/rev   ..

 Align seq 1/1  to reverse of: AI791640  from: 1  to: 464

       1 MetLysGlnArgThrSerPro 7
         |||||||||||||||||||||
     195 ATGAAGCAAAGAACAAGCCCT 175

Seq_name: qb_est70:BE073809

Seq_documentation_block:
LOCUS        BE073809       497 bp    mRNA            EST       09-JUN-2000
DEFINITION  hc09b10561-230200-034-e04 B10561 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE073809
```

```
VERSION     BE073809.1  GI:8421143
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 497)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski Almeida,S., Briones,M.R.,
            Naqui,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo SP,
            Brazil
            Tel: +55 11 2704922
            Fax: +55 11 2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/cgi-bin/getlumi2.pl?r1 &t2 RC0 BT0561 240
            200-034 604&t3 2000 02 28&t4 1)
            Seq primer: puc 18 forward
            High quality sequence start: 49
            High quality sequence stop: 495.
FEATURES             Location/Qualifiers
     source          1..497
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="BT0561"
                     /dev_stage="Adult"
                     /note="Organ: breast; Vector: pucl8; Site 1: SmaI; Site 2:
                     SmaI; A mini library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716    Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT       98 a    158 c    149 g     92 t
ORIGIN


alignment_scores:
             Quality:    7.00              Length:       7
               Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2 COPY_215_222 x BE073809/rc 2    ..

 Align seq 1/1  to reverse of: BE073809  from: 1  to: 497

        2 LysGluArgThrSerProGly 8
          ||||||||||||||||||||||
      258 AAGGAACGGACGAGCCCTGGG 238
```

```
>M cl: US-09-126-945B-2_COPY_229_237 to: GenEmbl:*   out_format : pir

Date: Sep 14, 2001  9:39 PM

About: Results were produced by the GenCore software, version 4.5.
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL framer_p2n.model -DEV xlp
 -Q "a:\_2385[1]\sq_1_1026[2]\45\pat_1[1].fil" -DB GenEmbl -QFMT fastap -SUFFIX=olip2n.rge -GAPOP 4.500
 -GAPEXT 0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT 60.000
 -PGAPOP 6.000 -PGAPEXT 7.000 -YGAPOP=60.000 -YGAPEXT 60.000
 -DELOP 6.000 -DELEXT 7.000 -START=1 -MATRIX olig0
 -TRANS human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE quality
 -THR_MIN=1 -ALIGN 15 -MODE LOCAL -OUTFMT pir  NORM ext  MINLEN 0
 -MAXLEN=2000000000  USER=US09126945_#CGN1_1_7594 -NCPU 6
 -ICPU 3 -LONGLOG -NO_XLPXY  WAIT -THREADS 1

Search information block:
Query: US-09-126-945B-2_COPY_229_237
Query length: 9
Database: GenEmbl:*
Database sequences: 1444157
Database length: -856060004
Search time (sec): 6503.700000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000


score_list:
Sequence              Strd Orig    ZScore    EScore  Len  ! Documentation ..
qb_pat1:AX048148        +     9.00  169.45    0.3349    1894  ! AX048148 Sequence 1 from Patent
qb_pr4:AF071538         +     9.00  169.45    0.3349    1894  ! AF071538 Homo sapiens Ets tran
qb_pr1:AB031549         +     9.00  169.43    0.3354    1858  ! AB031549 Homo sapiens ESE mRNA
qb_pat1:AX048159        +     9.00  166.35    0.4979    3317  ! AX048159 Sequence 12 from Pate
qb_pr6:AL157372         -     9.00  145.30    7.41      150660  ! AL157372 Human DNA sequence
qb_htg4:AC012625        +     7.00  107.01    1.0e-03   147363  ! AC012625 Homo sapiens chrome
qb_htg2:AC010264        +     7.00  106.54    1.1e-03   149428  ! AC010264 Homo sapiens chrome
qb_htg23:AL590453       -     7.00  106.23    1.1e-03   158056  ! AL590453 Homo sapiens chrome
qb_htg5:AC015815        -     7.00  106.11    1.1e-03   161549  ! AC015815 Homo sapiens clone
qb_htg14:AC068169       +     7.00  105.49    1.2e-03   180717  ! AC068169 Homo sapiens chrome
qb_htg8:AC022535        -     7.00  105.49    1.2e-03   180766  ! AC022535 Homo sapiens chrome
qb_htg6:AC018893        -     7.00  105.47    1.2e-03   181371  ! AC018893 Homo sapiens chrome
qb_htg3:AC012297        +     7.00  105.47    1.2e-03   181455  ! AC012297 Mus musculus clone
qb_htg14:AC068595       +     7.00  105.46    1.2e-03   181790  ! AC068595 Homo sapiens chrome
qb_pr2:AC005740         -     7.00  105.31    1.3e-03   186780  ! AC005740 Homo sapiens chrome
qb_htg13:AC048338       +     7.00  105.31    1.3e-03   186785  ! AC048338 Homo sapiens chrome
qb_pat1:AX033426        +     6.00  128.24    66.09     87  ! AX033426 Sequence 7 from Patent
qb_ba1:AF128789         +     6.00  124.45    107.49    173  ! AF128789 Uncultured bacterium
qb_sts2:KIAA0438        +     6.00  123.90    115.29    171  ! AJ229436 Kluyveromyces lactis
qb_ba2:BSPR2192         +     6.00  123.40    122.88    209  ! X89362 Bacterial sp. gene for
qb_pat1:AX033420        +     6.00  123.07    128.24    222  ! AX033420 Sequence 1 from Patent
qb_pat1:AX033421        +     6.00  123.07    128.24    222  ! AX033421 Sequence 2 from Patent
qb_vi2:MVVLTR28         -     6.00  122.71    134.42    237  ! Z31634 Maedi visna virus (clone
qb_ba2:AF261047         +     6.00  120.56    176.99    350  ! AF261047 Caulobacter sp. SKA25
qb_sts2:G49284          +     6.00  120.28    183.49    368  ! G49284 stSK92978_1117 chromosom
qb_sts2:G17829          +     6.00  119.82    194.53    400  ! G17829 human STS SHGC 14910 clo
qb_sts2:G52172          +     6.00  119.42    204.75    430  ! G52172 SHGC 79378 Human Homo sa
qb_ba1:AF025565         +     6.00  119.37    206.10    434  ! AF025565 Caulobacter sp. MED22
qb_pr1:AB032698         -     6.00  119.14    212.41    453  ! AB032698 Homo sapiens gene for
qb_ba1:AF010010         +     6.00  118.93    218.05    470  ! AF010010 Unidentified enbacteri
qb_sts2:G62461          +     6.00  118.85    220.35    477  ! G62461 SHGC-140042 Human Homo s
qb_pr7:CDC2L1S08        +     6.00  118.82    221.33    480  ! AF080678 Homo sapiens PITSLRE p
qb_ba2:AF186448         +     6.00  118.73    223.93    488  ! AF186448 Uncultured sponge symb
qb_pl1:AE074894         +     6.00  118.46    231.67    512  ! AE074894 Oryza sativa clone BGA
qb_pl2:AF227001         +     6.00  118.46    231.67    512  ! AF227001 Oryza sativa subsp. ja
qb_pl1:AE032696         +     6.00  118.45    231.99    513  ! AE032696 Oryza sativa NBG ERK t
qb_ba2:AF223464         -     6.00  118.41    233.27    517  ! AF223464 Synechococcus PCC7943
qb_pl3:OSHB55812        +     6.00  118.41    233.27    517  ! Y09811 O.sativa L. putative dis
qb_pl3:PAU75965         +     6.00  118.15    241.20    542  ! U75965 Pichia antillensis 26S r
qb_pat2:AX077285        -     6.00  116.67    243.71    550  ! AX077285 Sequence 9 from Pat
qb_htg12:AC041718       -     6.00  117.91    248.71    566  ! AC041718 Giardia intestinali
qb_pl2:AF413465         +     6.00  117.72    254.59    585  ! AF413465 Sporopachydermia sp
qb_pr7:CDC21S08         +     6.00  117.66    256.74    592  ! AF080689 Homo sapiens protei
qb_sts2:G52697          +     6.00  117.56    260.10    603  ! G52697 SHGC-79508 Human Homo
qb_pat1:AR099058        +     6.00  117.48    262.85    612  ! AR099058 Sequence 146 from p

seq_name: qb_pat1:AX048148

seq_documentation_block:
LOCUS       AX048148      1894 bp    DNA             PAT       15-DEC-2000
DEFINITION  Sequence 1 from Patent WO0070092.
ACCESSION   AX048148
VERSION     AX048148.1  GI:11876967
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1894)
  AUTHORS   Kaufmann,J. and Xin,H.
  TITLE     Expression of ets-domain proteins in cancer
  JOURNAL   Patent: WO 0070092-A 1 23-NOV-2000;
            Chiron Corporation (US)
FEATURES             Location/Qualifiers
     source          1..1894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             416..1423
                     /note="human GKETS."
                     /codon_start=1
                     /protein_id="CAC19002.1"
                     /db_xref="GI:11876968"
                     /translation="MGSASPGLSSVSPSHLLLPFPIVSKIGLEKAAAGAVGLERLWS
                     PSPEAIIEQGLSAEYLSYFDMLYPPDSSWAAKAGASSKPPPPPPEGPEVIDSLAPA
                     GSLDEVDQSLTFFFHSLPQVGMVVCEVLPDIFLACFLLNCIALPMPWSPSNVLRWLL
                     WLRLQLRLPMGKAEQELLAGKELCAMSEEQFRQRSFLGQDVELHAHLDLWKSAAWMKER
                     LSPGATRYCASLSFLSWLDSFVDSSCSQQPLHLWQFLKELLLKRGSYGFFLPWLNKFK
                     GIFFIFLSALVAPLWCLKFNFPAMNSTDFSPSLPQYAYEFCLLPKLLDLSQRLVYQFVHP
                     L"
BASE COUNT       368 a    654 c    571 g    302 t
ORIGIN


alignment_scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_229_237 x AX048148   ..

 Align seq 1/1  to: AX048148  from: 1  to: 1894

        1 SerThrSerGlnGluSerIrpThrAsp 9
          |||||||||||||||||||||||||||
     1160 TCGACCAGTCAGGAGAGCTCGACCGAC 1126

seq_name: qb_pr4:AF071538

seq_documentation_block:
LOCUS       AF071538      1894 bp    mRNA            PRI       10-JAN-2000
DEFINITION  Homo sapiens Ets transcription factor PDEF (PDEF) mRNA, complete
            cds.
ACCESSION   AF071538
VERSION     AF071538.1  GI:4007417
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1894)
  AUTHORS   Oettgen,P., Finger,E., Sun,Z., Akbarali,Y., Thamrongsak,U.,
```

```
            Bolton,J., Grall,F., Dube,A., Weiss,A., Brown,L., Quinn,G., Kas,K.,
            Endress,G., Kunsch,C. and Libermann,T.A.
  TITLE     PDEF, a novel prostate epithelium specific ets transcription
            factor, interacts with the androgen receptor and activates
            prostate specific antigen gene expression
  JOURNAL   J. Biol. Chem. 275 (2), 1215-1225 (2000)
  MEDLINE   20092894
REFERENCE   2  (bases 1 to 1894)
  AUTHORS   Libermann,T.A., Oettgen,P., Kunsch,C. and Finger,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUN-1998) Medicine, Beth Israel Deaconess Medical
            Center, 330 Brookline Ave., Boston, MA 02215, USA
FEATURES             Location/Qualifiers
     source          1..1894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6p21.3"
                     /tissue_type="prostate"
     gene            1..1894
                     /gene="PDEF"
     CDS             416..1423
                     /gene="PDEF"
                     /note="prostate epithelium specific"
                     /codon_start=1
                     /product="Ets transcription factor PDEF"
                     /protein_id="AAC95296.1"
                     /db_xref="gi:4007418"
                     /translation="MGSASPGLSSVSPSHLLLPPDIVSRIGLEKAAAGAVGLERRDWS
                     PSPPATPEQGLSAFYLSYFQMLYPEDSSWAARAPGASSPEEPPPFPQCPVIDSQAPA
                     GSLDLVPGGLTLEPHSLEQVQSMVVGPVLKDIETACKLLNITADPMDWSPSNVQKWLL
                     WTEHQYRLFPMGKAFQELAGKELCAMSEEQFRQWRSRLQGDVLHAHLLDWKSAAWMKER
                     TSPGATHYCASISEESWLDSFVDSSCSGQPTHLWQFLKELLLKPHSYGRFIRWLNKEK
                     GIFKIEDSAQVAPLWGTPKNRPAMNYDKLSRSLPQYYKVGIIERDIDTSQRLVYQFVHP
                     I"
BASE COUNT       368 a    653 c    571 g    302 t
ORIGIN


alignment scores:
            Quality:     9.00              Length:       9
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2 COPY 229-237 x AF071538  ..

 Align seq 1/1  to: AF071538  from: 1  to: 1894

        1 SerThrSerGluGluSerIrpThrAsp 9
          |||||||||||||||||||||||||||
     1100 TCGACCAGTGAGGAGAGCTGGACCGAC 1126

seq_name: qb_pri:AB031549

seq_documentation_block:
LOCUS        AB031549      1898 bp    mRNA            PRI       20-JAN-2000
DEFINITION   Homo sapiens PSE mRNA for prostate ets, complete cds.
ACCESSION    AB031549
VERSION      AB031549.1  GI:6721497
KEYWORDS     prostate ets.
SOURCE       Homo sapiens adult male prostate epithelium prostate cancer
             cell line:PC-3 cDNA to mRNA.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (sites)
  AUTHORS    Yamada,N., Tamai,Y., Miyamoto,H. and Nozaki,M.
  TITLE      Cloning and expression of the mouse Pse gene encoding a novel Ets
             family member
  JOURNAL    Gene 241 (2), 267-274 (2000)
  MEDLINE    20137504
* REFERENCE   2  (bases 1 to 1898)
  AUTHORS    Nozaki,M.
  TITLE      Direct Submission
  JOURNAL    Submitted (25-APR-1999) to the DDBJ/EMBL/GenBank databases. Masami
             Nozaki, Osaka University, Institute for Microbial Diseases, 3-1,
             Yamadaoka, Suita, Osaka 565-0871, Japan
             (E-mail:mnozaki@biken.osaka-u.ac.jp, Tel:81-6-879-8338,
             Fax:81-6-879-8339)
FEATURES             Location/Qualifiers
     source          1..1898
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="PC-3"
                     /cell_type="prostate cancer"
                     /chromosome="6"
                     /dev_stage="adult"
                     /map="6p21.2-21.3"
                     /sex="male"
                     /tissue_type="prostate epithelium"
                     /note="vector:pBC18"
     gene            406..1413
                     /gene="PSE"
     CDS             406..1413
                     /gene="PSE"
                     /codon_start=1
                     /product="prostate ets"
                     /protein_id="BAA89543.1"
                     /db_xref="GI:6721498"
                     /translation="MGSASPGLSSVSPSHLLLPPDIVSRIGLEKAAAGAVGLERRDWS
                     PSPPATPEQGLSAFYLSYFQMLYPEDSSWAARAPGASSPEEPPPFPQCPVIDSQAPA
                     GSLDLVPGGLTLEPHSLEQVQSMVVGPVLKDIETACKLLNITADPMDWSPSNVQKWLL
                     WTEHQYRLFPMGKAFQELAGKELCAMSEEQFRQWRSRLQGDVLHAHLLDWKSAAWMKER
                     TSPGATHYCASISEESWLDSFVDSSCSGQPTHLWQFLKELLLKPHSYGRFIRWLNKEK
                     GIFKIEDSAQVAPLWGTPKNRPAMNYDKLSRSLPQYYKVGIIERDIDTSQRLVYQFVHP
                     I"
BASE COUNT       380 a    651 c    567 g    300 t
ORIGIN


alignment scores:
            Quality:     9.00              Length:       9
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2 COPY 229-237 x AB031549  ..

 Align seq 1/1  to: AB031549  from: 1  to: 1898

        1 SerThrSerGluGluSerIrpThrAsp 9
          |||| ||||||||||||||||||||||||
     1090 TCGACCAGTGAGGAGAGCTGGACCGAC 1116

seq_name: qb_pat1:AX048159

seq_documentation_block:
LOCUS        AX048159      3317 bp    DNA             PAT       15-DEC-2000
DEFINITION   Sequence 12 from Patent WO0070092.
ACCESSION    AX048159
VERSION      AX048159.1  GI:11876975
KEYWORDS     .
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 3317)
  AUTHORS    Kaufmann,Y. and Xin,H.
  TITLE      Expression of ets-domain proteins in cancer
  JOURNAL    Patent: WO 0070092-A 12 23-NOV-2000;
             Chiron Corporation (US)
FEATURES             Location/Qualifiers
     source          1..3317
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

```
                     /note "GSEF encoding sequence with promoter (Plus, 2A 2B)"
     promoter        1. .1380
                     /note "Promoter"
     CDS             1815. .2822
                     /note "GSEF coding sequence"
                     /codon_start 1
                     /protein_id "CAC19691.1"
                     /db_xref "GI:11876076"
                     /translation "MESAIP [illegible] PSFPAIPEQGLSAFYLSYFIMLYFFLSSWAAKAPGASSPEEPTEFPEQTPVIHSQAPA
                     GSIDIVPGGLTIEPHSLFQVQSMVVCEVLKDGTLACELLNCTALTMCWSPSNVSKWLL
                     WIEHQYRLPPMGFAFSFIACFFCAMCTSIFSFSKCVDLDAHLYWKSAANMKFR
                     TSPGATHYCASISEESWIDSEVDSSSSQPIHLWQFLKELLKPHSYCKFIPWLNKFK
                     SIFKTEDSAQVAKIWGTKENFLAMNTFLFSPRRNLSIVTLPVHF
                     I"
BASE COUNT       710 a    1026 c     970 g     711 t
ORIGIN


alignment scores:
            Quality:     9.00              Length:       9
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_229_237 x AX048159  ..

 Align seq 1/1  to: AX048159  from: 1  to: 3417

        1 SerThrSerGluGluSerTrpThrAsp 9
          |||||||||||||||||||||||||||
     2499 TCGACCAGTGAGGAGAGCTGGACCGAC 2525

seq_name: qb_pr6:AL157372

seq_documentation_block:
LOCUS        AL157372    150660 bp    DNA             PRI       29-NOV-2000
DEFINITION   Human DNA sequence from clone RP11-375E1 on chromosome 6, complete
             sequence.
ACCESSION    AL157372
VERSION      AL157372.18  GI:11493236
KEYWORDS     HTG.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 150660)
  AUTHORS    Skuce,C.
  TITLE      Direct Submission
  JOURNAL    Submitted (29-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
             CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
             requests: clonerequest@sanger.ac.uk
COMMENT      On Nov 30, 2000 this sequence version replaced gi:11412772.
             During sequence assembly data is computed from overlapping clones.
             Where differences are found these are annotated as variations
             together with a note of the overlapping clone name. Note that the
             variation annotation may not be found in the sequence submission
             corresponding to the overlapping clone, as we submit sequences with
             only a small overlap as described above.
             This sequence has been finished according to sequence map criteria
             as follows. An attempt is made to resolve all sequencing problems,
             such as compressions and repeats, but not necessarily within known
             annotated repeat sequence elements. Where the sequence is
             ambiguous, there is an annotation using the 'unsure' feature key.
             The following abbreviations are used to associate primary accession
             numbers given in the feature table with their source databases:
             Em:, EMBL; Sw:, SWISSPROT; Tr:, TrEMBL; Wp:, WORMPEP; Information
             on the WORMPEP database can be found at
             http://www.sanger.ac.uk/Projects/C_elegans/wormpep. Supplied
             was generated from part of bacterial clone contigs of human
             chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
             Group. Further information can be found at
             http://www.sanger.ac.uk/HGP/Chr6/
             RP11-375E1 is from the library RPCI-11.2 constructed at the Roswell
             Park Cancer Institute by the group of Pieter de Jong. For further
             details see http://www.chori.org/bacpac/rpci_human_bac_library.htm
             VECTOR: pBACe3.6
             IMPORTANT: This sequence is not the entire insert of clone
             RP11-375E1 it may be shorter because we sequence overlapping
             sections only, except for a 100 base overlap.
             The true left end of clone RP11-375E1 is at 1 in this sequence. The
             true left end of clone RP4-391C22 is at 150561 in this sequence.
             The true right end of clone RP11-202L10 is at 11044 in this
             sequence.
FEATURES             Location/Qualifiers
     source          1. .150660
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /chromosome "6"
                     /clone "RP11-375E1"
                     /clone_lib "RPCI-11.2"
BASE COUNT     3542 a   3882 c   3740 g   3756 t
ORIGIN


alignment scores:
            Quality:     9.00              Length:       9
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_229_237 x AL157372/rev  ..

 Align seq 1/1  to reverse of: AL157372  from: 1  to: 150660

        1 SerThrSerGluGluSerTrpThrAsp 9
          |||||||||||||||||||||||||||
   124679 TCGACCAGTGAGGAGAGCTGGACCGAC 124653

seq_name: qb_htg4:AC012625

seq_documentation_block:
LOCUS        AC012625    137363 bp    DNA             HTG       05-MAY-2000
DEFINITION   Homo sapiens chromosome 5 clone CTD-2091N23, WORKING DRAFT
             SEQUENCE, 20 unordered pieces.
ACCESSION    AC012625
VERSION      AC012625.4  GI:7710258
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 137363)
  AUTHORS    DOE Joint Genome Institute.
  TITLE      Sequencing of Human Chromosome 5
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 137363)
  AUTHORS    DOE Joint Genome Institute.
  TITLE      Direct Submission
  JOURNAL    Submitted (31-OCT-1999) Production Sequencing Facility, DOE Joint
             Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT      On May 5, 2000 this sequence version replaced gi:3441566.
             Genome Center
             Center: Joint Genome Institute
             Center Code: JGI
             Web site: http://www.jgi.doe.gov
             Project Information
             Center Project Name: 652828
             Center clone name: CTD-2091N23
             Quality Statistics
             Consensus quality: 108649 bases at least Q40
             Consensus quality: 124503 bases at least Q30
             Consensus quality: 127445 bases at least Q20
             Phrap Contig Info: 1.99000; pulse field gel estimation
```

```
            Estimated insert size: 145463; sum of contigs estimation
            Quality coverage:  3.67 in Q20 bases; pulse field gel estimation
            Quality coverage:  4.04 in Q20 bases; sum of contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 20 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1      2038: contig of 2038 bp in length
            *      2039      2138: gap of unknown length
            *      2139      4146: contig of 2008 bp in length
            *      4147      4246: gap of unknown length
            *      4247      7425: contig of 3179 bp in length
            *      7426      7525: gap of unknown length
            *      7526     10252: contig of 2727 bp in length
            *     10253     10352: gap of unknown length
            *     10353     13027: contig of 2675 bp in length
            *     13028     13127: gap of unknown length
            *     13128     16823: contig of 3696 bp in length
            *     16824     16923: gap of unknown length
            *     16924     20298: contig of 3375 bp in length
            *     20299     20398: gap of unknown length
            *     20399     25001: contig of 4603 bp in length
            *     25002     25101: gap of unknown length
            *     25102     28697: contig of 3596 bp in length
            *     28698     28797: gap of unknown length
            *     28798     33791: contig of 4994 bp in length
            *     33792     33891: gap of unknown length
            *     33892     39676: contig of 5785 bp in length
            *     39677     39776: gap of unknown length
            *     39777     44891: contig of 5115 bp in length
            *     44892     44991: gap of unknown length
            *     44992     50282: contig of 5291 bp in length
            *     50283     50382: gap of unknown length
            *     50383     55805: contig of 5423 bp in length
            *     55806     55905: gap of unknown length
            *     55906     62175: contig of 6270 bp in length
            *     62176     62275: gap of unknown length
            *     62276     69875: contig of 7600 bp in length
            *     69876     69975: gap of unknown length
            *     69976     81873: contig of 11898 bp in length
            *     81874     81973: gap of unknown length
            *     81974     95425: contig of 13452 bp in length
            *     95426     95525: gap of unknown length
            *     95526    113632: contig of 18107 bp in length
            *    113633    113732: gap of unknown length
            *    113733    137363: contig of 23631 bp in length.
FEATURES             Location/Qualifiers
     source          1..137463
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /chromosome "5"
                     /clone "CTD-2091N23"
                     /clone_lib "CalTech human BAC library B"
BASE COUNT    37115 a   30453 c   30565 g   37212 t    2018 others
ORIGIN


alignment scores:
            Quality:    7.00              Length:       7
              Ratio:  1.000               Gaps:         0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-9458-2_COPY_229_237 x AC012625  ..

 Align seq 1/1  to: AC012625  from: 1  to: 137463

       1 SerThrSerGluGluSerTrp 7
         ||||||||||||||||||||||

    93337 TCAACCTCAGAGGAGAGTTGG 93357

seq_name: gb_htg2:AC010264

seq_documentation_block:
LOCUS        AC010264    149428 bp    DNA              HTG       18-JUL-2000
DEFINITION   Homo sapiens chromosome 5 clone CTC-468K18, WORKING DRAFT SEQUENCE,
             23 ordered pieces.
ACCESSION    AC010264
VERSION      AC010264.5  GI:9256174
KEYWORDS     HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 149428)
  AUTHORS    DOE Joint Genome Institute.
  TITLE      Sequencing of Human Chromosome 5
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 149428)
  AUTHORS    DOE Joint Genome Institute.
  TITLE      Direct Submission
  JOURNAL    Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
             Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT      On Jul 18, 2000 this sequence version replaced gi:7710802.
             ~~~~~Genome Center
             Center: Joint Genome Institute
             Center Code: JGI
             Web site: http://www.jgi.doe.gov
             ~~~~~~~
             Project Information
             Center Project Name: 465150
             Center clone name: CTC-HSPC_468K18
             ~~~~~~
             Summary Statistics
             Consensus quality: 136437 bases at least Q40
             Consensus quality: 144306 bases at least Q30
             Consensus quality: 146065 bases at least Q20
             Estimated insert size: 150000; pulse field gel estimation
             Estimated insert size: 148478; sum-of-contigs estimation
             Quality coverage:  4.59 in Q20 bases; pulse field gel estimation
             Quality coverage:  4.64 in Q20 bases; sum-of-contigs estimation.
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 23 contigs. Gaps between the contigs
             * are represented as runs of N. The order of the pieces
             * is believed to be correct as given, however the sizes
             * of the gaps between them are based on estimates that have
             * provided by the submittor.
             * This sequence will be replaced
             * by the finished sequence as soon as it is available and
             * the accession number will be preserved.
             *         1      8196: contig of 8196 bp in length
             *      8197      8296: gap of unknown length
             *      8297     12916: contig of 4620 bp in length
             *     12917     13016: gap of unknown length
             *     13017     14756: contig of 1740 bp in length
             *     14757     14856: gap of unknown length
             *     14857     19143: contig of 4287 bp in length
             *     19144     19243: gap of unknown length
             *     19244     20796: contig of 1553 bp in length
             *     20797     20896: gap of unknown length
             *     20897     28283: contig of 7387 bp in length
             *     28284     28383: gap of unknown length
             *     28384     43450: contig of 15067 bp in length
             *     43451     43550: gap of unknown length
             *     43551     55287: contig of 11737 bp in length
             *     55288     55387: gap of unknown length
             *     55388     73601: contig of 18214 bp in length
             *     73602     73701: gap of unknown length
             *     73702     78295: contig of 4594 bp in length
             *     78296     78395: gap of unknown length
             *     78396     81832: contig of 3437 bp in length
             *     81833     81932: gap of unknown length
```

```
            *    81934    83998: contig of 2065 bp in length
            *    83999    84098: gap of unknown length
            *    84099    86531: contig of 2433 bp in length
            *    86532    86631: gap of unknown length
            *    86632    88757: contig of 2126 bp in length
            *    88758    88857: gap of unknown length
            *    88858    95081: contig of 6224 bp in length
            *    95082    95181: gap of unknown length
            *    95182    98288: contig of 3107 bp in length
            *    98289    98388: gap of unknown length
            *    98389   107715: contig of 9327 bp in length
            *   107716   107815: gap of unknown length
            *   107816   112037: contig of 4222 bp in length
            *   112038   112137: gap of unknown length
            *   112138   114033: contig of 1896 bp in length
            *   114034   114133: gap of unknown length
            *   114134   119274: contig of 5141 bp in length
            *   119275   119374: gap of unknown length
            *   119375   126879: contig of 7505 bp in length
            *   126880   126979: gap of unknown length
            *   126980   129104: contig of 2125 bp in length
            *   129105   129204: gap of unknown length
            *   129205   149428: contig of 20224 bp in length.
FEATURES             Location/Qualifiers
     source          1. .149428
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /chromosome "6"
                     /clone "CIC-468K18"
                     /clone_lib "Caltech human BAC library C"
BASE COUNT     39562 a   32858 c   33087 g   43817 t    2204 others
ORIGIN


alignment_scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_229_237 x AC010264    ..

 Align seq 1/1  to: AC010264  from: 1  to: 149428

       1 SerPhrSerGluGluSerIrp 7
         |||||||||||||||||||||
   58621 TCAACCTCAGAGGAGAGTGG 58641

seq_name: qb_htq23:AL590453

seq_documentation_block:
LOCUS        AL590453   158056 bp    DNA              HTG       11-APR-2001
DEFINITION   Homo sapiens chromosome 1 clone RP11 294K17, *** SEQUENCING IN
             PROGRESS ***, 42 unordered pieces.
ACCESSION    AL590453
VERSION      AL590453.2  GI:13568475
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 158056)
  AUTHORS    Melay,K.
  TITLE      Direct Submission
  JOURNAL    Submitted (10-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
             CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
             requests: clonerequest@sanger.ac.uk
COMMENT      On Apr 9, 2001 this sequence version replaced gi:13561040.
             -------------- Genome Center
             Center: Sanger Centre
             Center code: SC
             Web site: http://www.sanger.ac.uk
             Contact: humquery@sanger.ac.uk

             ------ Project Information
Center project name: bA294K17
             ------------ Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: DyeTerminator Big Dye; 100% of reads
Consensus quality: 142549 bases at least Q40
Consensus quality: 149195 bases at least Q30
Consensus quality: 152697 bases at least Q20
Insert size: 154956; sum of contigs
Insert size: 185496; 5.1% error; agarose tp
Quality coverage: 3.19x in Q20 bases; sum of contigs quality
coverage: 3.00x in Q20 bases; agarose tp
             ---------- ----.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 42 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     5502: contig of 5502 bp in length
*     5503  5602: gap of      100 bp
*     5603     7659: contig of 2057 bp in length
*     7660  7759: gap of      100 bp
*     7760    14575: contig of 6816 bp in length
*    14576 14675: gap of      100 bp
*    14676    16874: contig of 2199 bp in length
*    16875 16974: gap of      100 bp
*    16975    21650: contig of 4676 bp in length
*    21651 21750: gap of      100 bp
*    21751    34805: contig of 13055 bp in length
*    34806 34905: gap of      100 bp
*    34906    39111: contig of 4206 bp in length
*    39112 39211: gap of      100 bp
*    39212    45997: contig of 6786 bp in length
*    45998 46097: gap of      100 bp
*    46098    51125: contig of 5028 bp in length
*    51126 51225: gap of      100 bp
*    51226    56031: contig of 4806 bp in length
*    56032 56131: gap of      100 bp
*    56132    58915: contig of 2784 bp in length
*    58916 59015: gap of      100 bp
*    59016    63228: contig of 4213 bp in length
*    63229 63328: gap of      100 bp
*    63329    71780: contig of 8452 bp in length
*    71781 71880: gap of      100 bp
*    71881    75164: contig of 3284 bp in length
*    75165 75264: gap of      100 bp
*    75265    78091: contig of 2827 bp in length
*    78092 78191: gap of      100 bp
*    78192    81844: contig of 3653 bp in length
*    81845 81944: gap of      100 bp
*    81945    84340: contig of 2396 bp in length
*    84341 84440: gap of      100 bp
*    84441    87361: contig of 2921 bp in length
*    87362 87461: gap of      100 bp
*    87462    92210: contig of 4749 bp in length
*    92211 92310: gap of      100 bp
*    92311   104325: contig of 12015 bp in length
*   104326 104425: gap of      100 bp
*   104426   108093: contig of 3668 bp in length
*   108094 108193: gap of      100 bp
*   108194   111477: contig of 3284 bp in length
*   111478 111577: gap of      100 bp
*   111578   113965: contig of 2388 bp in length
*   113966 114065: gap of      100 bp
*   114066   119354: contig of 5289 bp in length
*   119355 119454: gap of      100 bp
*   119455   125386: contig of 5932 bp in length
*   125387 125486: gap of      100 bp
*   125487   129401: contig of 3915 bp in length
```

```
            *    129402 129501: gap of       100 bp
            *    129502   132234: contig of 2733 bp in length
            *    132235 132334: gap of       100 bp
            *    132335   136298: contig of 3964 bp in length
            *    136299 136398: gap of       100 bp
            *    136399   140650: contig of 4252 bp in length
            *    140651 140750: gap of       100 bp
            *    140751   143027: contig of 2277 bp in length
            *    143028 143127: gap of       100 bp
            *    143128   147928: contig of 4801 bp in length
            *    147929 148028: gap of       100 bp
            *    148029   158056: contig of 10028 bp in length.
FEATURES             Location/Qualifiers
     source          1..158056
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /chromosome "1"
                     /clone "RP11 294K17"
                     /clone_lib "RPCI-11.2"
     misc_feature    1..5502
                     /note "assembly_fragment:01564
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    5603..7659
                     /note "assembly_fragment:01319
                     fragment_chain:1"
     misc_feature    7760..14575
                     /note "assembly_fragment:00389
                     fragment_chain:2"
     misc_feature    14676..16874
                     /note "assembly_fragment:00208
                     fragment_chain:2"
     misc_feature    16975..21650
                     /note "assembly_fragment:00597
                     fragment_chain:2"
     misc_feature    21751..34805
                     /note "assembly_fragment:00070
                     fragment_chain:3"
     misc_feature    34906..38911
                     /note "assembly_fragment:00620
                     fragment_chain:3"
     misc_feature    39212..45997
                     /note "assembly_fragment:00432
                     fragment_chain:4"
     misc_feature    46098..51125
                     /note "assembly_fragment:00209
                     fragment_chain:4"
     misc_feature    51226..56041
                     /note "assembly_fragment:00467
                     fragment_chain:5"
     misc_feature    56142..58915
                     /note "assembly_fragment:00552
                     fragment_chain:5"
     misc_feature    59016..63228
                     /note "assembly_fragment:00523
                     fragment_chain:6"
     misc_feature    63329..71780
                     /note "assembly_fragment:01155
                     fragment_chain:6"
     misc_feature    71881..75164
                     /note "assembly_fragment:00526
                     fragment_chain:7"
     misc_feature    75265..78091
                     /note "assembly_fragment:00435
                     fragment_chain:7"
     misc_feature    78192..81844
                     /note "assembly_fragment:00676
                     fragment_chain:8"
     misc_feature    81945..84340
                     /note "assembly_fragment:00059
                     fragment_chain:8"
*    misc_feature    84441..87461
                     /note "assembly_fragment:00892
                     fragment_chain:9"
     misc_feature    87462..92210
                     /note "assembly_fragment:01109
                     fragment_chain:9"
     misc_feature    92311..104425
                     /note "assembly_fragment:01244
                     fragment_chain:10"
     misc_feature    104426..108093
                     /note "assembly_fragment:00876
                     fragment_chain:10"
     misc_feature    108194..111477
                     /note "assembly_fragment:01305
                     fragment_chain:11"
     misc_feature    111578..113965
                     /note "assembly_fragment:00912
                     fragment_chain:11"
     misc_feature    114066..119354
                     /note "assembly_fragment:01452
                     fragment_chain:12"
     misc_feature    119455..125386
                     /note "assembly_fragment:01527
                     fragment_chain:12"
     misc_feature    125487..129401
                     /note "assembly_fragment:00726"
     misc_feature    129502..132234
                     /note "assembly_fragment:01169"
     misc_feature    132335..136298
                     /note "assembly_fragment:01184"
     misc_feature    136399..140650
                     /note "assembly_fragment:01230"
     misc_feature    140751..143027

alignment scores:
            quality:     7.00              length:       7
              Ratio:    1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2-COPY-229-237 x AL590453/rev  ..

 Align seq 1/1  to reverse of: AL590453  from: 1  to: 158056

        2 ThrSerGlnGluSerIrpThr 8
          |||| ||||||||||||||||
  123731 ACAAGTGAAGAATCCTGGACT 123711

seq_name: gb:htg5:AC015815

seq_documentation_block:
LOCUS        AC015815    161549 bp    DNA              HTG       30-MAR-2000
DEFINITION   Homo sapiens clone RP11 21015, WORKING DRAFT SEQUENCE, 21 unordered
             pieces.
ACCESSION    AC015815
VERSION      AC015815.4  GI:7441971
KEYWORDS     HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 161549)
  AUTHORS    Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE      Homo sapiens, clone RP11 21015
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 161549)
  AUTHORS    Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
             Baldwin,J., Barna,N., Beckerly,R., Boqustavkiy,L., Boukhalter,B.,
             Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
             Cooke,P., DeAreliano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
             Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
             Galagan,J., Gardyna,S., Grant,G., Baqos,B., Hearford,A., Horton,L.,
             Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
             Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
```

```
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., McPherson,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Testave,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 29, 2000 this sequence version replaced gi:5070069.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ------------ Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            ------------ Project Information
                Center project name: L4073
                Center clone name: 21_0_15
            ------ ------ Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 151333 bases at least Q40
                Consensus quality: 155764 bases at least Q30
                Consensus quality: 157650 bases at least Q20
                Insert size: 157000; agarose-tp
                Insert size: 159549; sum-of-contigs
                Quality coverage: 4.9 in Q20 bases; agarose-tp
                Quality coverage: 4.8 in Q20 bases; sum-of-contigs
            ------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 21 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1138: contig of 1138 bp in length
            *     1139  1238: gap of      100 bp
            *     1239     2713: contig of 1475 bp in length
            *     2714  2813: gap of      100 bp
            *     2814     4948: contig of 2135 bp in length
            *     4949  5048: gap of      100 bp
            *     5049     7939: contig of 2891 bp in length
            *     7940  8039: gap of      100 bp
            *     8040     9902: contig of 1863 bp in length
            *     9903 10002: gap of      100 bp
            *    10003    12571: contig of 2569 bp in length
            *    12572 12671: gap of      100 bp
            *    12672    16792: contig of 4121 bp in length
            *    16793 16892: gap of      100 bp
            *    16893    21320: contig of 4428 bp in length
            *    21321 21420: gap of      100 bp
            *    21421    28865: contig of 7445 bp in length
            *    28866 28965: gap of      100 bp
            *    28966    34699: contig of 5734 bp in length
            *    34700 34799: gap of      100 bp
            *    34800    40251: contig of 5452 bp in length
            *    40252 40351: gap of      100 bp
            *    40352    47897: contig of 7546 bp in length
            *    47898 47997: gap of      100 bp
            *    47998    55766: contig of 7769 bp in length
            *    55767 55866: gap of      100 bp
            *    55867    63968: contig of 8102 bp in length
            *    63969 64068: gap of      100 bp
            *    64069    74694: contig of 10626 bp in length
            *    74695 74794: gap of      100 bp
            *    74795    84737: contig of 9943 bp in length
            *    84738 84837: gap of      100 bp
            *    84838    98420: contig of 13583 bp in length
            *    98421 98420: gap of      100 bp
            *    98421   110611: contig of 12191 bp in length
            *   110612 110711: gap of      100 bp
            *   110712   122418: contig of 11707 bp in length
            *   122419 122518: gap of      100 bp
            *   122519   137004: contig of 14486 bp in length
            *   137005 137104: gap of      100 bp
            *   137105   161549: contig of 24445 bp in length
FEATURES             Location/Qualifiers
     source          1..161549
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11 21o15"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..1138
                     /note="assembly_fragment"
     misc_feature    1239..2713
                     /note="assembly_fragment"
     misc_feature    2814..4948
                     /note="assembly_fragment"
     misc_feature    5049..7939
                     /note="assembly_fragment"
     misc_feature    8040..9902
                     /note="assembly_fragment"
     misc_feature    10003..12571
                     /note="assembly_fragment"
     misc_feature    12672..16792
                     /note="assembly_fragment"
     misc_feature    16893..21420
                     /note="assembly_fragment"
     misc_feature    21421..28865
                     /note="assembly_fragment"
     misc_feature    28866..34699
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    34800..40251
                     /note="assembly_fragment"
     misc_feature    40352..47897
                     /note="assembly_fragment"
     misc_feature    47998..55766
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    55867..63968
                     /note="assembly_fragment"
     misc_feature    64069..74694
                     /note="assembly_fragment"
     misc_feature    74795..84737
                     /note="assembly_fragment"
     misc_feature    84838..98420
                     /note="assembly_fragment"
     misc_feature    98421..110611
                     /note="assembly_fragment"
     misc_feature    110712..122418
                     /note="assembly_fragment"
     misc_feature    122519..137004
                     /note="assembly_fragment"
     misc_feature    137105..161549
                     /note="assembly_fragment"
BASE COUNT     41638 a   38819 c   39363 g   39720 t    2009 others
ORIGIN


alignment_scores:
              Quality:    7.00              Length:       7
                Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_229_237 x AC015815/rev   1..
```

```
 Align seq 1/1  to reverse of: AC015815  from: 1  to: 161549

       1 SerThrSerGluGluSerTrp 7
         !!!!!!!!!!!!!!!!!!!!
  124433 AGCACCTCTCAAGAATCTGG 124414

seq name: gb htg14:AC068169

seq_documentation block:
LOCUS       AC068169    180717 bp    DNA              HTG       15 NOV 2000
DEFINITION  Homo Sapiens chromosome 17 clone RP11 651B2 map 17, WORKING DRAFT
            SEQUENCE, 15 unordered pieces.
ACCESSION   AC068169
VERSION     AC068169.3  GI:11136767
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 180717)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11 651B2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 180717)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,B., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Boda,F.,
            Boqusklavkiy,L., Bouklejalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRoccpe,K., Lamazares,R., Landers,P., Lehoczky,J.,
            Levine,R., Lieu,C., Lin,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mompa,V., Morrow,I.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,I., O'Donnell,P.,
            O'Neil,D., Olivar,I.M., Oliver,J., Petersen,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rojez,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Triqilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,R., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR 2000) Whitehead Institute/MIT Center for Genome
            Research, 420 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 10, 2000 this sequence version replaced gi:8389515.
            All repeats were identified using RepeatMasker
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
                    -- Genome Center
                Center: Whitehead Institute/MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www.genome.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
                    -- Project Information
                Center project name: LL0247
                Center clone name: 651 B 2
                    -- Summary Statistics
                Sequencing vector: M13; M77815; 43% of reads
                Sequencing vector: Plasmid; n/a; 57% of reads
                Chemistry: Dye terminator BigDye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 174839 bases at least Q40
                Consensus quality: 177110 bases at least Q30
                Consensus quality: 178528 bases at least Q20
                Insert size: 179000; agarose fp
                Insert size: 179417; sum of contigs
                Quality coverage: 8.5 in Q20 bases; agarose fp
                Quality coverage: 8.9 in Q20 b
              * NOTE: This is a 'working draft' sequence. It currently
              * consists of 15 contigs. The true order of the pieces
              * is not known and their order in this sequence record is
              * arbitrary. Gaps between the contigs are represented as
              * runs of N, but the exact sizes of the gaps are unknown.
              * This record will be updated with the finished sequence
              * as soon as it is available and the accession number will
              * be preserved.
              *       1      756: contig of 756 bp in length
              *     757    856: gap of      100 bp
              *     857     2208: contig of 1352 bp in length
              *    2209    2308: gap of      100 bp
              *    2309     4715: contig of 2407 bp in length
              *    4716    3815: gap of      100 bp
              *    3816     5977: contig of 2162 bp in length
              *    5978    6077: gap of      100 bp
              *    6078    57570: contig of 51493 bp in length
              *   57571   57670: gap of      100 bp
              *   57671    60754: contig of 3084 bp in length
              *   60755   60854: gap of      100 bp
              *   60855    65276: contig of 4422 bp in length
              *   65277   65376: gap of      100 bp
              *   65377    71222: contig of 5846 bp in length
              *   71223   71322: gap of      100 bp
              *   71323    81741: contig of 10419 bp in length
              *   81742   81841: gap of      100 bp
              *   81842    91954: contig of 10113 bp in length
              *   91955   92054: gap of      100 bp
              *   92055    104839: contig of 12785 bp in length
              *  104840  104939: gap of      100 bp
              *  104940    125196: contig of 20257 bp in length
              *  125197  125296: gap of      100 bp
              *  125297    149973: contig of 24677 bp in length
              *  149974  150073: gap of      100 bp
              *  150074    178960: contig of 28887 bp in length
              *  178961  179060: gap of      100 bp
              *  179061    180717: contig of 1657 bp in length.
FEATURES             Location/Qualifiers
     source          1. .180717
                     /organism="Homo sapiens"
                     /db_xref "taxon:9606"
                     /chromosome="17"
                     /map "17"
                     /clone "RP11-651B2"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .756
                     /note "assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    857. .2208
                     /note "assembly_fragment"
     misc_feature    2309. .4715
                     /note "assembly_fragment"
     misc_feature    3816. .5977
                     /note "assembly_fragment"
     misc_feature    6078. .57570
                     /note "assembly_fragment"
     misc_feature    57671. .60754
                     /note "assembly_fragment"
     misc_feature    60855. .65276
                     /note "assembly_fragment"
     misc_feature    65377. .71222
                     /note "assembly_fragment"
     misc_feature    71323. .81741
                     /note "assembly_fragment"
     misc_feature    81842. .91954
                     /note "assembly_fragment"
     misc_feature    92055. .104839
                     /note "assembly_fragment"
     misc_feature    104940. .125196
                     /note "assembly_fragment"
     misc_feature    125297. .149973
                     /note "assembly_fragment"
     misc_feature    150074. .178960
```

```
                     /note "assembly_fragment"
     misc_feature    179061..180717
                     /note "assembly_fragment
                     clone_end:17"
                     vector_side:right"
BASE COUNT    48215 a  44298 c  41765 g  46002 t    1407 others
ORIGIN


alignment_scores:
            Quality:    7.00             Length:       7
              Ratio:  1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_229_237 & AC022535/rev  ..

 Align seq 1/1  to reverse of AC022535  from: 1  to: 180766

       1 SerThrSerGluGluSerIrp 7
         |||||||||||||||||||||
  111957 AGCACCICTGAAGAATCTTGG 111937

seq_name: gb_htg8:AC022535

seq_documentation_block:
LOCUS        AC022535   180766 bp    DNA              HTG       17 FEB 2001
DEFINITION   Homo sapiens chromosome 10 clone RP11-344C15, WORKING DRAFT
             SEQUENCE, 7 unordered pieces.
ACCESSION    AC022535
VERSION      AC022535.4  GI:12957674
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 180766)
  AUTHORS    Smith,D.R.
  TITLE      Genome Therapeutics Corporation Sequencing Center: Human Genome
             Sequence Data
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 180766)
  AUTHORS    Smith,D.R.
  TITLE      Direct Submission
  JOURNAL    Submitted (06 FEB 2000) Genome Therapeutics Corporation, 100 Beaver
             Street, Waltham, MA 02453, USA
COMMENT      On Feb 17, 2001 this sequence version replaced gi:9929856.
             -------------- Genome Center
                            Center: Genome Therapeutics Corporation
                            Center code: GTC
                            Web site: http://www.genomecorp.com/
                            Contact: gtc_sequencer-genomecorp.com
                        ------------ Project Information
                            Center project name: hq163
                        -------------- Summary Statistics
                            Sequencing vector: N/A
                            Chemistry: Dye-terminator BigDye; 100% of reads
                            Assembly program: Phrap; version 960415
                            Consensus quality: 263547 bases at least Q40
                            Consensus quality: 275045 bases at least Q30
                            Consensus quality: 280559 bases at least Q20
                            Insert size: 180166; sum of contigs
                            Quality coverage: 3.7x in Q20 bases; sum-of-contigs
                        ---------- End ----------
                * NOTE: This is a 'working draft' sequence. It currently
                * consists of 7 contigs. The true order of the pieces
                * is not known and their order in this sequence record is
                * arbitrary. Gaps between the contigs are represented as
                * runs of N, but the exact sizes of the gaps are unknown.
                * This record will be updated with the finished sequence
                * as soon as it is available and the accession number will
                * be preserved.
                *        1    52202: contig of 52202 bp in length
                *    52203    52302: gap of unknown length
                *    52303    64092: contig of 11790 bp in length
                *    64093    64192: gap of unknown length
                *    64193    97297: contig of 33105 bp in length
                *    97298    97397: gap of unknown length
                *    97398   119643: contig of 22246 bp in length
                *   119644   119743: gap of unknown length
                *   119744   139719: contig of 19976 bp in length
                *   139720   139819: gap of unknown length
                *   139820   146932: contig of 7113 bp in length
                *   146933   147032: gap of unknown length
                *   147033   180766: contig of 33734 bp in length
FEATURES             Location/Qualifiers
     source          1..180766
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /chromosome="10"
                     /clone="RP11-344C15"
                     /clone_lib="RPCI-11"
     misc_feature    1..52202
                     /note "assembly_name:Contig162
                     clone end:17"
     misc_feature    52303..64092
                     /note "assembly_name:Contig157"
     misc_feature    64193..97297
                     /note "assembly_name:Contig161"
     misc_feature    97398..119643
                     /note "assembly_name:Contig158"
     misc_feature    119744..139719
                     /note "assembly_name:Contig159"
     misc_feature    139820..146932
                     /note "assembly_name:Contig156"
     misc_feature    147033..180766
                     /note "assembly_name:Contig160
                     clone_end:SP6"
BASE COUNT    55646 a  33727 c  34075 g  56715 t    603 others
ORIGIN


alignment_scores:
            Quality:    7.00             Length:       7
              Ratio:  1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_229_237 & AC022535/rev  ..

 Align seq 1/1  to reverse of AC022535  from: 1  to: 180766

       2 IhrSerGluGluSerIrpIhr 8
         ||||||||||||||||||||||
   64355 AGTAGTGAGGAGAGTTGGACA 64335

seq_name: gb_htg6:AC018893

seq_documentation_block:
LOCUS        AC018893   181371 bp    DNA              HTG       16 MAR 2001
DEFINITION   Homo sapiens chromosome 2 clone RP11-514F21, WORKING DRAFT
             SEQUENCE, 8 unordered pieces.
ACCESSION    AC018893
VERSION      AC018893.4  GI:13357548
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 181371)
  AUTHORS    Waterston,R.H.
  TITLE      The sequence of Homo sapiens clone
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 181371)
  AUTHORS    Waterston,R.H.
  TITLE      Direct Submission
```

```
  JOURNAL   Submitted (21 DEC 1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Mar 16, 2001 this sequence version replaced gi:11192247.

                          Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/est.cgi
                    Project Information
            Center project name: B_NH0514F21
                      Summary Statistics
            Sequencing vector: M13; 57%
            Sequencing vector: plasmid; 39%
            Chemistry: Dye primer ET; 57% of reads
            Chemistry: Dye terminator Big Dye; 39% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 179264 bases at least Q40
            Consensus quality: 179908 bases at least Q30
            Consensus quality: 180430 bases at least Q20
            Insert size: 167000; agarose tp
            Insert size: 181171; sum of contigs
            Quality coverage: 7.71 in Q20 bases; agarose tp
            Quality coverage: 7.13 in Q20 bases; sum of contigs
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   145816: contig of 145816 bp in length
            *  145817   145916: gap of unknown length
            *  145917   149800: contig of 3884 bp in length
            *  149801   149900: gap of unknown length
            *  149901   181471: contig of 31471 bp in length.
FEATURES             Location/Qualifiers
     source          1..181471
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="RP11-514F21"
     misc_feature    1..145816
                     /note="assembly_name:Contig10"
     misc_feature    145917..149800
                     /note="assembly_name:Contig8"
     misc_feature    149901..181471
                     /note="assembly_name:Contig9"
BASE COUNT    48265 a  43845 c  42897 g  46164 t    200 others
ORIGIN


alignment scores:
            Quality:     7.00              Length:       7
              Ratio:    1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_229_237 x AC018893(rev

 Align seq 1/1  to reverse of: AC018893  from: 1  to: 181471

       1 SerThrSerGluGluSerTrp 7
         |||||||||||||||||||||
  144490 TCAACCTCTGAGGAGTCCTGG 144470

seq name: gb|htg3|AC012297

seq documentation_block:
LOCUS       AC012297    181455 bp    DNA             HTG       27-JAN-2001
DEFINITION  Mus musculus clone RP23-78H9, WORKING DRAFT SEQUENCE, 9 unordered
            pieces.
ACCESSION   AC012297
VERSION     AC012297.4  GI:12583840
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 181455)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-78H9
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 181455)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boqustavkiy,I., Bonkhalter,B.,
            Brown,A., Castle,A., Colanqelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., Fitzhuqh,W., Forrest,C., Funke,R., Gage,D.,
            Galaqan,J., Gardyna,S., Grant,G., Haqos,B., Beaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lebaczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., Meslark,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pellara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stande Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 420 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jan 27, 2001 this sequence version replaced gi:2230064.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
                          Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
                ---------  Project Information
                Center project name: L4207
                Center clone name: 78_H_9
                      Summary Statistics
                Sequencing vector: M13; M77815; 55% of reads
                Sequencing vector: Plasmid; n/a; 45% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 178104 bases at least Q40
                Consensus quality: 179410 bases at least Q30
                Consensus quality: 180022 bases at least Q20
                Insert size: 182000; agarose-tp
                Insert size: 180655; sum-of contigs
                Quality coverage: 11.2 in Q20 bases; agarose tp
                Quality coverage: 11.4 in Q20 b.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   36507: contig of 36507 bp in length
            *   36508   36607: gap of      100 bp
            *   36608   38030: contig of 1423 bp in length
            *   38031   38130: gap of      100 bp
            *   38131   40636: contig of 2506 bp in length
            *   40637   40736: gap of      100 bp
            *   40737   49410: contig of 8674 bp in length
            *   49411   49510: gap of      100 bp
            *   49511   64625: contig of 15115 bp in length
            *   64626   64725: gap of      100 bp
            *   64726   87916: contig of 23191 bp in length
```

```
             *    87917 88016: gap of      100 bp
             *    88017   107007: contig of 18991 bp in length
             *   107008 107107: gap of      100 bp
             *   107108   130410: contig of 23303 bp in length
             *   130411 130510: gap of      100 bp
             *   130511   181455: contig of 50946 bp in length.
FEATURES             Location/Qualifiers
     source          1. .181455
                     /organism "Mus musculus"
                     /db_xref="taxon:10090"
                     /clone "RP23-78H9"
                     /clone_lib "RPCI-23 Female Mouse BAC"
     misc_feature    1. .36507
                     /note "assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    36608. .38030
                     /note="assembly_fragment"
     misc_feature    38131. .40646
                     /note "assembly_fragment"
     misc_feature    40737. .49410
                     /note "assembly_fragment"
     misc_feature    49511. .64025
                     /note "assembly_fragment"
     misc_feature    64726. .87916
                     /note "assembly_fragment"
     misc_feature    88017. .107007
                     /note "assembly_fragment"
     misc_feature    107108. .130410
                     /note="assembly_fragment"
     misc_feature    130511. .181455
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    49553 a  43935 c  42125 g  45525 t    817 others
ORIGIN


alignment_scores:
            Quality:    7.00            Length:      7
              Ratio:   1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_229_237 x AC012297   .

 Align seq 1/1  to: AC012297  from: 1  to: 181455

       1 SerThrSerGluGluSerTrp 7
         ||||||||||||||||||||
   51665 TCAACAAGTGAAGAGAGCTGG 51685

seq_name: gb_htg14:AC068595

seq_documentation_block:
LOCUS       AC068595   181790 bp    DNA             HTG       07-JUL-2000
DEFINITION  Homo sapiens chromosome 14 clone RPII-233E12, WORKING DRAFT
            SEQUENCE, 14 unordered pieces.
ACCESSION   AC068595
VERSION     AC068595.1  GI:7705227
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 181790)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 181790)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT
            ------- Genome Center -------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web [illegible]
            ------- Project Information -------
            Center project name: B_NH0233E12
            ------- Summary Statistics -------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 175488 bases at least Q40
            Consensus quality: 177190 bases at least Q30
            Consensus quality: 178285 bases at least Q20
            Insert size: 170000; agarose-1p
            Insert size: 180490; sum-of-contigs
            Quality coverage: 4.84 in Q20 bases; agarose 1p
            Quality coverage: 4.60 in Q20 bases; sum of contigs
            ------- -----  -------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      1269: contig of 1269 bp in length
            *     1270      1369: gap of unknown length
            *     1370      2538: contig of 1169 bp in length
            *     2539      2638: gap of unknown length
            *     2639      3754: contig of 1116 bp in length
            *     3755      3854: gap of unknown length
            *     3855      5884: contig of 2030 bp in length
            *     5885      5984: gap of unknown length
            *     5985      9254: contig of 3270 bp in length
            *     9255      9354: gap of unknown length
            *     9355     14017: contig of 4663 bp in length
            *    14018     14117: gap of unknown length
            *    14118     26183: contig of 12066 bp in length
            *    26184     26283: gap of unknown length
            *    26284     37688: contig of 11405 bp in length
            *    37689     37788: gap of unknown length
            *    37789     50284: contig of 12496 bp in length
            *    50285     50484: gap of unknown length
            *    50485     65926: contig of 15542 bp in length
            *    65927     66026: gap of unknown length
            *    66027     83741: contig of 17715 bp in length
            *    83742     83841: gap of unknown length
            *    83842    111890: contig of 28049 bp in length
            *   111891    111990: gap of unknown length
            *   111991    138461: contig of 26471 bp in length
            *   138462    138561: gap of unknown length
            *   138562    181790: contig of 43229 bp in length.
FEATURES             Location/Qualifiers
     source          1. .181790
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /chromosome "14"
                     /clone "RPII 233E12"
     misc_feature    1. .1269
                     /note "assembly_name:Contig44"
     misc_feature    1370. .2538
                     /note "assembly_name:Contig46"
     misc_feature    2639. .3754
                     /note "assembly_name:Contig40"
     misc_feature    3855. .5884
                     /note "assembly_name:Contig44"
     misc_feature    5985. .9254
```

```
                     /note "assembly_name:Contig45"
     misc_feature    9455..14017
                     /note "assembly_name:Contig46"
     misc_feature    14118..26183
                     /note "assembly_name:Contig47"
     misc_feature    26284..37688
                     /note "assembly_name:Contig48"
     misc_feature    37789..50284
                     /note "assembly_name:Contig49
                     clone_end:SP6
                     vector_side:right"
     misc_feature    50385..65926
                     /note "assembly_name:Contig50"
     misc_feature    66027..83741
                     /note "assembly_name:Contig51"
     misc_feature    83842..111890
                     /note "assembly_name:Contig52"
     misc_feature    111991..138461
                     /note "assembly_name:Contig53
                     clone_end:T7
                     vector_side:left"
     misc_feature    138562..181790
                     /note "assembly_name:Contig54"
BASE COUNT    60021 a   42260 c   42074 g   56131 t    1304 others
ORIGIN


Alignment scores:
             Quality:     7.00              Length:       7
               Ratio:    1.000               Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

Alignment block:
 US-09-126-945B-2_COPY_229_237 x AC068595


 Align seq 1/1   to: AC068595   from: 1   to: 181790

        2 ThrSerGloGluSerTrpThr 8
          ||||||||||||||||||||
    50076 ACAACTGAAGAATCCTGGACT 50096

seq name: gb_pr2:AC005740

seq documentation block:
LOCUS       AC005740    186780 bp    DNA              PRI       01-OCT-1998
DEFINITION  Homo sapiens chromosome 5p, BAC clone 50q21 (LBNL H154), complete
            sequence.
ACCESSION   AC005740
VERSION     AC005740.1  GI:4687210
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 186780)
  AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Cunning,K.M.,
            Kadner,K., Miquel,T., Miller,C., Pitluck,S., Pollard,M.,
            Rojeski,M., Subramanian,S. and Martin,C.H.
  TITLE     Sequencing of human chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 186780)
  AUTHORS   Bieke,D.D.
  TITLE     Large Scale Sequence Analysis and Annotation with the Sequence
            Comparison Analysis (SCAN) System
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 186780)
  AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Cunning,K.M.,
            Kadner,K., Miquel,T., Miller,C., Pitluck,S., Pollard,M.,
            Rojeski,M., Subramanian,S. and Martin,C.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-1998) Human Genome Center, DOE Joint Genome
            Institute, Lawrence Berkeley National Laboratory, MS 74-157,
            Berkeley, CA 94720, U.S.A.
COMMENT     Sequence submitted by:
            DOE Joint Genome Institute.
FEATURES             Location/Qualifiers
     source          1..186780
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="5p"
                     /clone="50q21"
                     /chromosome="5"
                     /note="LBNL H154"
     repeat_region   complement(1..74)
                     /rpt_family="Alu"
     repeat_region   868..1134
                     /rpt_family="Alu"
     repeat_region   complement(1208..1514)
                     /rpt_family="Alu"
     repeat_region   complement(3486..3634)
                     /rpt_family="L1"
     repeat_region   3648..3938
                     /rpt_family="Alu"
     repeat_region   3921..3943
                     /note="(A)23"
                     /rpt_type=tandem
                     /rpt_unit=A
     repeat_region   complement(5195..5272)
                     /rpt_family="MIR2"
     repeat_region   complement(5719..6005)
                     /rpt_family="Alu"
     repeat_region   complement(6469..6647)
                     /rpt_family="L1"
     repeat_region   complement(6688..6955)
                     /rpt_family="Alu"
     repeat_region   6821..6847
                     /note="(T)27"
                     /rpt_type=tandem
                     /rpt_unit=T
     repeat_region   complement(7611..7742)
                     /rpt_family="MLFa"
     repeat_region   complement(7833..7932)
                     /rpt_family="MSTc"
     repeat_region   8492..8601
                     /rpt_family="Alu"
     repeat_region   complement(9403..9676)
                     /rpt_family="Alu"
     repeat_region   complement(10100..10257)
                     /rpt_family="Alu"
     misc_feature    11124..11167
                     /note="GRAIL_2 exon:Dist exon, frame 1"
     repeat_region   11224..11247
                     /note="(ATTC)6"
                     /rpt_type=tandem
                     /rpt_unit=ATTC
     repeat_region   12044..12360
                     /rpt_family="Alu"
     repeat_region   12443..12469
                     /note="(A)27"
                     /rpt_type=tandem
                     /rpt_unit=A
     repeat_region   complement(12872..12811)
                     /rpt_family="L1"
     repeat_region   12900..13016
                     /rpt_family="Alu"
     repeat_region   complement(13084..13169)
                     /rpt_family="L1"
     repeat_region   complement(13597..13889)
                     /rpt_family="Alu"
     repeat_region   complement(15617..15941)
                     /rpt_family="Alu"
     repeat_region   15793..15817
                     /note="(T)25"
                     /rpt_type=tandem
                     /rpt_unit=T
     repeat_region   complement(16185..16462)
```

```
                     /rpt_family "Alu"
     repeat_region   complement(16604. .16902)
                     /rpt_family="Alu"
     repeat_region   17036. .17206
                     /rpt_family="MIT1"
     repeat_region   17220. .17832
                     /rpt_family "Alu"
     repeat_region   17626. .17649
                     /note="(A)24"
                     /rpt_type=tandem
                     /rpt_unit=A
     repeat_region   17874. .18004
                     /rpt_family "MIT1f"
     repeat_region   18255. .18529
                     /rpt_family="Alu"
     repeat_region   complement(18691. .18870)
                     /rpt_family="MER3"
     repeat_region   complement(18788. .18822)
                     /rpt_family="MER33"
     repeat_region   complement(18983. .19265)
                     /rpt_family "Alu"
     repeat_region   20976. .21209
                     /rpt_family="Alu"
     repeat_region   21298. .21414
                     /rpt_family "L1"
     repeat_region   complement(21413. .21577)
                     /rpt_family="Alu"
     repeat_region   complement(21730. .22011)
                     /rpt_family "Alu"
     repeat_region   22064. .22271
                     /rpt_family="Alu"
     misc_feature    22531. .22595
                     /note "GRAIL 2 exon, frame 1"
     repeat_region   22780. .23054
                     /rpt_family="Alu"
     repeat_region   23190. .23489
                     /rpt_family="Alu"
     repeat_region   23849. .24002
                     /rpt_family="MER39B"
     repeat_region   complement(24000. .26149)
                     /rpt_family="L1"
     repeat_region   26149. .27398
                     /rpt_family="L1"
     repeat_region   27432. .27706
                     /rpt_family="Alu"
     repeat_region   27721. .28129
                     /rpt_family "MER39"
     repeat_region   28498. .28774
                     /rpt_family="Alu"
     repeat_region   complement(29403. .29490)
                     /rpt_family="MIR2"
     repeat_region   29525. .29716
                     /rpt_family "Alu"
     repeat_region   30196. .30275
                     /rpt_family="MIR"
     repeat_region   complement(30377. .30662)
                     /rpt_family="Alu"
     repeat_region   complement(30672. .31137)
                     /rpt_family "L1"
     repeat_region   31138. .31564
                     /rpt_family="MER57B"
     repeat_region   complement(31565. .31687)
                     /rpt_family="L1"
     misc_feature    join(31640. .31733,31791. .32094)
                     /note="85% & 91% identity EST zb50104.r1"
                     /db_xref "dbEST:W24400"
     misc_feature    31917. .32270
                     /note="90% identity EST zi04c10.r1"
                     /db_xref "dbEST:AA009751"
     repeat_region   32354. .32655
                     /rpt_family="Alu"
     repeat_region   complement(32690. .32798)
                     /rpt_family="Alu"
     repeat_region   complement(32793. .32997)
                     /rpt_family "L1"
     repeat_region   33000. .33296
                     /rpt_family "Alu"
     repeat_region   34147. .34354
                     /rpt_family "Alu"
     repeat_region   complement(34383. .34822)
                     /rpt_family "L1"
     repeat_region   35142. .35273
                     /rpt_family="MER87"
     repeat_region   36021. .36304
                     /rpt_family "Alu"
     repeat_region   36482. .36594
                     /rpt_family="MIR"
     misc_feature    36830. .36949
                     /note "GRAIL 2 excellent exon, frame 1"
     repeat_region   complement(37800. .37874)
                     /rpt_family="MER91B"
     repeat_region   38633. .38910
                     /rpt_family="Alu"
     repeat_region   38909. .38945
                     /note="(AAT)9"
                     /rpt_type=tandem
                     /rpt_unit AAT
     repeat_region   complement(39679. .39973)

alignment_scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 "B.FG.1.7-93.P.2_7.FY_229_237 x AC005740.Dev   ..

 Align set 1/1  to reverse of: AC005740  from: 1  to: 186780

       1 SerThrSerGlnGluSerTrp 7
         |||||||||||||||||||||
  115877 TCAACCTCAGAGGAGAGTTGG 115857
```

```
OM of: US-09-126-945B-2_COPY_229_237.rge  N ?????? ; ???] *    ?? ?????  ???

Date: Sep 14, 2001, 9:48 PM

About: Results were produced by the GenCore software, version 4.5
       Copyright (c) 1993-2000 Compugen Ltd

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgm2_1/PSPUB/pubseq/pubseq2_5/45/runs/_1400_2001_14:51_5???_???/??PP_????? ????_1.???
 -DB=N_Geneseq_0601 -QFMT=fastap -SUFFIX=olip2n.rng -GAPOP 4.500
 -GAPEXT 0.050 -MINMATCH=0.100 -LOOPCL 0.000 -LOOPEXT 0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT 60.000
 -DELOP 6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE quality
 -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT pir -NORM=ext -MINLEN 0
 -MAXLEN 2000000000 -USER=US09126945_?CGN1_1_408 -NCPU 6 -ICPU=?
 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-126-945B-2_COPY_229_237
Query length: 9
Database: N_Geneseq_0601:*
Database sequences: 730101
Database length: 813950809
Search time (sec): 680.150000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT=60.000


score_list:
Sequence            Strd Orig    ZScore    EScore Len  ! Documentation  ..
/SIDSL/seqdata/geneseq/geneseqn/NA2000 DAT-AAA06617 +     6.00   183.73  0.0291     278  !
/SIDSL/seqdata/geneseq/geneseqn/NA2000 DAT-AAP22828 +     6.00   174.42  0.0959    1997  !
/SIDSL/seqdata/geneseq/geneseqn/NA2000 DAT-AA.50641 +     6.00   170.64  0.1556    1861  !
/SIDSL/seqdata/geneseq/geneseqn/NA2001 DAT-AAC93261 +     6.00   170.64  0.1556    1894  !
/SIDSL/seqdata/geneseq/geneseqn/NA2001 DAT-AAC93266 +     6.00   166.82  0.2541    3317  !
/SIDSL/seqdata/geneseq/geneseqn/NA2000 DAT-AAA73715 +     6.00   130.35  21.15      87  ! F
/SIDSL/seqdata/geneseq/geneseqn/NA2000 DAT-AAA73706 +     6.00   125.06  47.99      222  !
/SIDSL/seqdata/geneseq/geneseqn/NA2000 DAT-AAA73710 +     6.00   125.06  47.99      222  !
/SIDSL/seqdata/geneseq/geneseqn/NA2000 DAT-AAC58077 +     6.00   120.22  106.18     515  !
/SIDSL/seqdata/geneseq/geneseqn/NA2001 DAT-AAE50261 +     6.00   119.77  108.11     550  !
/SIDSL/seqdata/geneseq/geneseqn/NA1998 DAT-AAV04607 +     6.00   119.04  116.50     612  !
/SIDSL/seqdata/geneseq/geneseqn/NA2001 DAT-AAC90889 +     6.00   119.04  116.50     612  !
/SIDSL/seqdata/geneseq/geneseqn/NA2000 DAT-AAA73714 +     6.00   116.53  124.49     660  !
/SIDSL/seqdata/geneseq/geneseqn/NA2000 DAT-AAC93196 +     6.00   116.05  132.51     728  !
/SIDSL/seqdata/geneseq/geneseqn/NA2000 DAT-AAE.4167 +     6.00   117.92  134.30     720  !
/SIDSL/seqdata/geneseq/geneseqn/NA2000 DAT-AAA74805 +     6.00   117.92  134.30     720  !
/SIDSL/seqdata/geneseq/geneseqn/NA2000 DAT-AAC93241 +     6.00   117.92  221.97    1249  !
/SIDSL/seqdata/geneseq/geneseqn/NA2000 DAT-AAC77601 +     6.00   117.92  221.97    1249  !
/SIDSL/seqdata/geneseq/geneseqn/NA2000 DAT-AAC40807 +     6.00   117.29  241.00    1425  !
/SIDSL/seqdata/geneseq/geneseqn/NA2000 DAT-AAC52605 +     6.00   116.73  262.77    1561  !
/SIDSL/seqdata/geneseq/geneseqn/NA1998 DAT-AAV42711 +     6.00   112.41  276.36    1443  !
/SIDSL/seqdata/geneseq/geneseqn/NA1998 DAT-AAV72456 +     6.00   111.71  268.53    1735  !
/SIDSL/seqdata/geneseq/geneseqn/NA1999 DAT-AAY74013 +     6.00   111.19  327.92    1953  !
/SIDSL/seqdata/geneseq/geneseqn/NA1999 DAT-AAZ42120 +     6.00   111.19  319.30    1929  !
/SIDSL/seqdata/geneseq/geneseqn/NA2000 DAT-AAC77241 +     6.00   111.01  326.35    1987  !
/SIDSL/seqdata/geneseq/geneseqn/NA2000 DAT-AAC77846 +     6.00   110.96  298.50    2002  !
/SIDSL/seqdata/geneseq/geneseqn/NA2000 DAT-AAA74711 +     6.00   110.09  367.22    2274  !
/SIDSL/seqdata/geneseq/geneseqn/NA1998 DAT-AAV29148 +     6.00   109.62  350.32    2176  !
/SIDSL/seqdata/geneseq/geneseqn/NA2000 DAT-AAC77316 +     6.00   109.60  391.13    2414  !
/SIDSL/seqdata/geneseq/geneseqn/NA2000 DAT-AAA73711 +     6.00   109.53  394.91    2471  !
/SIDSL/seqdata/geneseq/geneseqn/NA2000 DAT-AAE15828 +     6.00   108.94  431.99    2738  !
/SIDSL/seqdata/geneseq/geneseqn/NA2000 DAT-AAA27464 +     6.00   105.82  634.31    4248  !
/SIDSL/seqdata/geneseq/geneseqn/NA1996 DAT-AAA19252 +     6.00   105.10  662.91    4167  !
/SIDSL/seqdata/geneseq/geneseqn/NA2000 DAT-AA.11.1. +     6.00   100.75  1.2e+03   8048  !
/SIDSL/seqdata/geneseq/geneseqn/NA2000 DAT-AAY85955 +     6.00   100.75  1.2e+03   8048  !
/SIDSL/seqdata/geneseq/geneseqn/NA2000 DAT-AAB20950 +     6.00    92.63  3.4e+03  20433  !
/SIDSL/seqdata/geneseq/geneseqn/NA2000 DAT-AAA74929 +     6.00    92.63  3.4e+03  20433  !
/SIDSL/seqdata/geneseq/geneseqn/NA1998 DAT-AAV22110 +     6.00    91.35  1.1e+03  35521  !
/SIDSL/seqdata/geneseq/geneseqn/NA2000 DAT-AAB21427 +     6.00    71.25  1.9e+04  209273 !
/SIDSL/seqdata/geneseq/geneseqn/NA2000 DAT-AAP22305 +     6.00    68.05  7.9e+04  1082
/SIDSL/seqdata/geneseq/geneseqn/NA2000 DAT-AAA86649 -     5.00   124.82  67.28      18
/SIDSL/seqdata/geneseq/geneseqn/NA2000 DAT-AAA86650 -     5.00   124.82  67.28      18
/SIDSL/seqdata/geneseq/geneseqn/NA2000 DAT-AAA86651 -     5.00   124.82  67.28      18
/SIDSL/seqdata/geneseq/geneseqn/NA2000 DAT-AAA39981 -     5.00   122.61  73.78      20
/SIDSL/seqdata/geneseq/geneseqn/NA2001 DAT-AAC67164 -     5.00   122.61  73.78      20

seq name: /SIDSL/seqdata/geneseq/geneseqn/NA2000 DAT-AAA06617

seq_documentation_block:
ID   AAA06617 standard; cDNA; 278 BP.
XX
AC   AAA06617;
XX
DT   13-JUN-2000  (first entry)
XX
DE   Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:398.
XX
KW   Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW   immunogenic; cytostatic; vaccine; ss.
XX
OS   Homo sapiens.
XX
PN   WO200004149-A2.
XX
PD   27-JAN-2000.
XX
PF   14-JUL-1999;   99WO US15838.
XX
PR   14-JUL-1998;   98US-0115453.
PR   14-JUL-1998;   98US-0116134.
PR   23-SEP-1998;   98US-0159812.
PR   23-SEP-1998;   98US-0159822.
PR   15-JAN-1999;   99US-0232149.
PR   15-JAN-1999;   99US-0232880.
PR   09-APR-1999;   99US-0288946.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Dillon DC,  Harlocker SL,  Yuqin J,  Xu J,  Mitcham JL;
XX
DR   WPI, 2000-171268/15.
XX
PT   New polypeptide useful for treating and diagnosing prostate cancer
PT   comprises an immunogenic portion of prostate tumor protein -
XX
PS   Claim 50; Page 239; 263pp; English.
XX
CC   The present invention describes isolated polypeptides, comprising an
CC   immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC   and polynucleotides encoding them have cytostatic activity and can be
CC   used in vaccines and in gene therapy. The polypeptides and
CC   polynucleotides encoding them, antibodies presenting cells which express
CC   the polypeptides, antibodies against the polypeptides and vaccines
CC   comprising them can be used for inhibiting the development of prostate
CC   cancer in a patient. The polypeptides can be used to generate antibodies
CC   or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC   the polynucleotides encoding the polypeptides can be used as a probe or
CC   to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CC   AAY82028 to AAY82029 represent sequences used in the exemplification of
CC   the present invention.
XX
SQ   Sequence 278 BP; 56 A; 85 C; 87 G; 49 T; 1 other;


alignment scores:
                Quality:    9.00              Length:      9
                  Ratio:   1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment file:
US-09-126-945B-2_COPY_229_237 x AAA06617  ..
```

```
 Align seq 1/1  to: AAA06617  from: 1  to: 278

        1 SerThrSerGluGluSerTrpThrAsp 9
          |||||||||||||||||||||||||||
      134 TCGACCAGTGAGGAGAGCTGGACCGAC 160

seq_name: /tmp1/pepdata/genseqp/genseqp_NA.tfa  AAF21828

seq_documentation_block:
ID   AAF21828 standard; DNA; 1087 BP.
XX
AC   AAF21828;
XX
DT   27-MAR-2001  (first entry)
XX
DE   Human breast and ovarian cancer associated antigen gene SEQ ID 215.
XX
KW   Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
KW   nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW   antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
KW   antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
KW   Addison's disease; allergy; autoimmune haemolytic anaemia;
KW   autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW   multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW   cardiovascular disorder; wound healing; neurological disease; ds.
XX
OS   Homo sapiens.
XX
PN   WO200055173-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05881.
XX
PR   12-MAR-1999;  99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-611515/58.
DR   P-PSDB; AAB68925.
XX
PT   New human breast and ovarian cancer associated gene sequences and the
PT   polypeptides encoded by these genes, useful in the prevention,
PT   treatment and diagnosis of cancer, immune disorders, cardiovascular
PT   disorders and neurological diseases
XX
PS   Claim 1; Page 642-643; 1299pp; English.
XX
CC   Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
CC   proteins AAB58711 - AAB59128. The DNA and protein sequences are
CC   associated with breast and ovarian cancer. Included in the invention are
CC   sequences AAF22032 - AAF22040 and AAB59129 which are used in the
CC   isolation and characterisation of the DNA and protein sequences of the
CC   invention. The breast and ovarian cancer associated DNA, protein, agonist
CC   or antagonist sequences exhibit cytostatic; immunosuppressive;
CC   nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
CC   antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
CC   antibacterial; antifungal; antiparasitic and cardiant activity. The
CC   polynucleotide and protein sequences are used in the diagnosis of cancer,
CC   particularly breast and ovarian cancer. The nucleic acid sequences,
CC   proteins, agonists and agonists may also be used in the diagnosis,
CC   prevention and treatment of immune disorders e.g. Addison's disease,
CC   allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC   diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC   arthritis and ulcerative colitis; cardiovascular disorders such as
CC   myocardial ischaemias; wound healing; neurological diseases such as
CC   cerebral anoxia and epilepsy; and infectious diseases.
XX
SQ   Sequence 1087 BP; 215 A; 347 C; 339 G; 182 T; 4 other;

alignment scores:
            Quality:    9.00              Length:      9
              Ratio:   1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 qy 1 tcgaccagtgaggagagctggaccgac 27
    |||||||||||||||||||||||||||
 AAF21828   561 tcgaccagtgaggagagctggaccgac 587

seq_name: /tmp1/pepdata/genseqp/genseqp_NA2000.tfa.AAA50691

seq_documentation_block:
ID   AAA50691 standard; cDNA; 1804 BP.
XX
AC   AAA50691;
XX
DT   28-MAY-2000  (first entry)
XX
DE   Human prostate derived Ets factor cDNA.
XX
KW   Prostate derived Ets factor; PDEF; human; chromosome 6p21.3; cancer;
KW   loss of heterozygosity; chromosomal translocation; linkage analysis;
KW   cytostatic; cardiant; immunosuppressive; cerebroprotective; fungicide;
KW   antibacterial; vulnerary; neuroprotective; antiparkinsonian; nootropic;
KW   anabolic; antiinflammatory; anorectic; hybridisation probe; forensic;
KW   tumour marker; diagnosis; treatment; prostate cancer; blood coagulation;
KW   autoimmune disorder; haematopoietic; immune/nervous system; stroke;
KW   neoplasm; microbial infection; tissue regeneration; heart attack;
KW   scarring; food additive; preservative; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             416..1423
FT                   /*tag= a
FT                   /product= "Human prostate derived Ets factor"
FT                   /note= "Expressed in prostate epithelium"
XX
PN   WO200006589-A1.
XX
PD   10-FEB-2000.
XX
PF   02-AUG-1999;  99WO-US17470.
XX
PR   31-JUL-1998;  98US-0126945.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (BETH ) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI   Libermann IA,  Oettgen JP,  Kunsch CA,  Endress GA,  Rosen CA;
XX
DR   WPI; 2000-195255/17.
DR   P-PSDB; AAY45003.
XX
PT   Novel prostate derived polypeptide, polynucleotide useful for
PT   diagnosis, prevention and treatment of prostate cancer, autoimmune
PT   disorders, microbial infections and also as food additive or
PT   preservative
XX
PS   Claim 4; Fig 1; 132pp; English.
XX
CC   The present cDNA sequence encodes a human prostate derived Ets factor
CC   (PDEF), isolated from human prostate carcinoma cDNA library
CC   (ATCC No.203072). PDEF gene expression is highest in tissues involved in
CC   androgen sensitivity, with enriched expression in prostate epithelium
CC   cells. PDEF gene is mapped to the human chromosome 6p21.3
CC   region that is associated with loss of heterozygosity and chromosomal
CC   translocations in various human cancers. PDEF has cytostatic, cardiant,
```

```
CC   immunosuppressive, cerebroprotective, fungicide, antibacterial,
CC   vulnerary, neuroprotective, antiparkinsonian, nootropic, anabolic,
CC   antiinflammatory and anorectic activity. fDEF polynucleotides are useful
CC   in linkage analysis as markers, as hybridisation probes for differential
CC   identification of the tissues or cell types and as polymorphic markers
CC   for forensic purposes. PDEF is useful as prostate-specific tumour marker
CC   for the diagnosis and treatment of prostate cancer. fDEF sequences are
CC   useful for treating autoimmune disorders, haematopoietic, blood
CC   coagulation, immune and nervous system disorders, hyperproliferative
CC   disorders like, neoplasms and microbial infections, heart attacks,
CC   stroke, scarring and for tissue regeneration. They are also useful as
CC   food additives or preservatives.
XX
SQ   Sequence 1894 BP; 468 A; 653 C; 571 G; 302 T; 0 other;


alignment_scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_229_237 x AAZ50691  ...

 Align_seq 1/1  to: AAZ50691  from: 1  to: 1894

        1 SerThrSerGluGluSerIrpThrAsp 9
          |||||||||||||||||||||||||||
     1160 TCGACCAGTGAGGAGAGCTGGACCGAC 1126

seq_name: /cf2s1/ftp-data/geneseq_prose_pub/AA...  ; A  AAZ50691

seq_documentation_block:
ID   AAZ50691 standard; DNA; 1894 BP.
XX
AC   AAZ50691;
XX
DT   16-MAR-2001  (first entry)
XX
DE   Gland-specific Ets transcription factor (GSEF) nucleotide sequence.
XX
KW   Transcription factor; gland specific Ets transcription factor; GSEF;
KW   metastatic potential; cancer; tumour; metastasis; breast; prostate;
KW   leukaemia; lymphoma; sarcoma; melanoma; chromosome 6p21.1 6p21.3; ss.
XX
OS   Homo sapiens.
XX
PN   WO200070092-A1.
XX
PD   23-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US13173.
XX
PR   14-MAY-1999;   99US-0134112.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Kaufmann J,  Xin H,  Harrowe G;
XX
DR   WPI; 2001-041019/05.
DR   P-PSDB; AAB49628.
XX
PT   Detecting metastatic and potential metastatic cancerous cells, useful
PT   for diagnosing, prognosing, grading and staging of cancers by detecting
PT   gland-specific Ets transcription factor gene product in a biological
PI   sample from a cell
XX
PS   Disclosure; Page 81-83; 95pp; English.
XX
CC   This invention relates to a method for the detection and determination of
CC   the metastatic potential of a cell. The method comprises detecting a
CC   gland-specific Ets transcription factor (GSEF) gene product in a test
CC   sample. Detection of a GSEF gene product in the test sample in amount
CC   lower than that in a normal cell is indicative of a cell with high
CC   metastatic potential. The method is useful for determining the metastatic
CC   potential of a cell, for the diagnosis and prognosis of cancer as well as
CC   grading and staging of cancers by detecting GSEF expression in a
CC   biological test sample. The method may also be used to monitor patients
CC   having a predisposition to develop a particular cancer. GSEF polypeptides
CC   are useful for producing antibodies, in cancer diagnosis, prognosis,
CC   grading, staging and management of breast and prostate tumours, and in
CC   detecting polymorphisms in the sequence. GSEF genes and proteins are also
CC   useful in gene therapy. GSEF gene product expression levels can be used
CC   in conjunction with any tissue in which an alteration in GSEF gene
CC   product expression levels is associated with development of a
CC   cancer-associated phenotype. Cancers, which can be monitored include
CC   cancers of the prostate, cervix, lung and colon, melanomas, colorectal
CC   adenocarcinomas, Wilm's tumour, retinoblastoma, sarcomas, myosarcomas,
CC   lung carcinomas, leukaemia, and lymphomas. The GSEF gene is located on
CC   human chromosome 6, specifically at 6p21.1-6p21.3. The present sequence
CC   represents the DNA encoding GSEF.
XX
SQ   Sequence 1894 BP; 468 A; 653 C; 571 G; 302 T; 0 other;


alignment_scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_229_237 x AAC83261  ...

 Align_seq 1/1  to: AAC83261  from: 1  to: 1894

        1 SerThrSerGluGluSerIrpThrAsp 9
          |||||||||||||||||||||||||||
     1160 TCGACCAGTGAGGAGAGCTGGACCGAC 1126

seq_name: /cf2s1/ftp-data/geneseq_prose_pub/AA...  ; A  AAC83261

seq_documentation_block:
ID   AAC83266 standard; DNA; 1917 BP.
XX
AC   AAC83266;
XX
DT   16-MAR-2001  (first entry)
XX
DE   Gland specific Ets transcription factor (GSEF) cDNA sequence.
XX
KW   Transcription factor; gland specific Ets transcription factor; GSEF;
KW   metastatic potential; cancer; tumour; metastasis; breast; prostate;
KW   leukaemia; lymphoma; sarcoma; melanoma; chromosome 6p21.1 6p21.3; ss.
XX
OS   Homo sapiens.
XX
PN   WO200070092-A1.
XX
PD   23-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US13173.
XX
PR   14-MAY-1999;   99US-0134112.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Kaufmann J,  Xin H,  Harrowe G;
XX
DR   WPI; 2001-041019/05.
DR   P-PSDB; AAB49633.
XX
PT   Detecting metastatic and potential metastatic cancerous cells, useful
PI   for diagnosing, prognosing, grading and staging of cancers by detecting
PT   gland-specific Ets transcription factor gene product in a biological
PT   sample from a cell
XX
```

PS Disclosure; Fig 2; 95pp; English.
XX
CC This invention relates to a method for the detection and determination of
CC the metastatic potential of a cell. The method comprises detecting a
CC gland specific ETS transcription factor (GSEF) gene product in a test
CC sample. Detection of a GSEF gene product in the test sample in amount
CC lower than that in a normal cell, is indicative of a cell with high
CC metastatic potential. The method is useful for determining the metastatic
CC potential of a cell, for the diagnosis and prognosis of cancer as well as
CC grading and staging of cancers by detecting GSEF expression in a
CC biological test sample. The method may also be used to monitor patients
CC having a predisposition to develop a particular cancer. GSEF polypeptides
CC are useful for producing antibodies, in cancer diagnosis, prognosis,
CC grading, staging and management of breast and prostate tumours, and in
CC detecting polymorphisms in the sequence. GSEF genes and proteins are also
CC useful in gene therapy. GSEF gene product expression levels can be used
CC in conjunction with any tissue in which an alteration in GSEF gene
CC product expression levels is associated with development of a
CC cancer associated phenotype. Cancers, which can be monitored include
CC cancers of the prostate, cervix, lung and colon, melanomas, colorectal
CC adenocarcinomas, Wilm's tumour, retinoblastoma, sarcomas, myosarcomas,
CC lung carcinomas, leukaemia, and lymphomas. The GSEF gene is located on
CC human chromosome 6, specifically at 6p21.1 6p21.3. The present sequence
CC represents the cDNA encoding GSEF.
XX
SQ Sequence 4317 BP; 710 A; 1026 C; 970 G; 611 T; 0 other;

alignment scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2 (COPY 229-237) x AAC83266

Align seq 1/1 to: AAC83266 from: 1 to: 4317

1 SerThrSerGluGluSerTrpThrAsp 9
|||||||||||||||||||||||||||
2499 TCGACCAGTGAGGAGAGCTGGACCGAC 2525

seq name: /SIDS2/seqdata/geneseq/geneseqp/NA/geneseqp_NA2000_SA1:AAC83266

seq documentation block:
ID AAC83266 standard; DNA; 4317 BP.
XX
AC AAC83266;
XX
DT 07-DEC-2000 (first entry)
XX
DE Functional fragment of human PITSLRE protein kinase gene IRES
XX
KW IRES; internal ribosome entry site; PITSLRE; human; gene therapy;
KW cancer; restenosis; p58; pl10; protein kinase; ds.
XX
OS Homo sapiens.
XX
PN WO200044896-A1.
XX
PD 03-AUG-2000.
XX
PF 26-JAN-2000; 2000WO-EP000643.
XX
PR 26-JAN-1999; 99EP-0200216.
XX
PA (VLAA) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOL.
XX
PI Cornelis S, Beyaert R;
XX
DR WPI; 2000-499331/44.
XX
PT Nucleic acids encoding internal ribosome entry sequences useful for
PT directing protein expression in gene therapy procedures
XX
PS Claim 5; Page 44; 57pp; English.
XX
CC The PITSLRE protein kinase gene can be translated to give two isoforms,
CC p110 and p58. Transcription of p58 mRNA was found to start at an
CC internal ribosome entry site (IRES). The IRES element was found to be
CC cell cycle regulated. p58 is produced during the G2/M stage of the cell
CC cycle. The present sequence is the functional fragment of the PITSLRE
CC protein kinase IRES. The IRES sequence and fragments of it may be used
CC to induce cell cycle dependent initiation of translation in eukaryotic
CC cells. Vectors containing the IRES may be used for the preparation of
CC compositions for the treatment of and/or prevention of a disease by
CC gene therapy. Such diseases may be cancer and restenosis.
XX
SQ Sequence 87 BP; 29 A; 17 C; 15 G; 27 T; 0 other;

alignment scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2 (COPY 229-237) x AAA73715

Align seq 1/1 to: AAA73715 from: 1 to: 87

1 SerThrSerGluGluSer 6
|||||||||||||||||||
40 AGCACCAGTGAGGAGTCA 57

seq name: /SIDS2/seqdata/geneseq/geneseqp/NA2000_SA1:AAA73709

seq documentation block:
ID AAA73709 standard; DNA; 222 BP.
XX
AC AAA73709;
XX
DT 07-DEC-2000 (first entry)
XX
DE Human PITSLRE protein kinase gene internal ribosome entry site.
XX
KW IRES; internal ribosome entry site; PITSLRE; human; gene therapy;
KW cancer; restenosis; p58; pl10; protein kinase; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT misc_signal 1..222
FT /*tag = a
FT /label= IRES
FT /note= "internal ribosome entry site"
XX
PN WO200044896-A1.
XX
PD 03-AUG-2000.
XX
PF 26-JAN-2000; 2000WO-EP000643.
XX
PR 26-JAN-1999; 99EP-0200216.
XX
PA (VLAA) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOL.
XX
PI Cornelis S, Beyaert R;
XX
DR WPI; 2000-499331/44.
XX
PT Nucleic acids encoding internal ribosome entry sequences useful for
PT directing protein expression in gene therapy procedures
XX
PS Claim 4; Page 42; 57pp; English.
XX

CC The PITSLRE protein kinase gene can be translated to give two isoforms, CC p110 and p58. Transcription of p58 mRNA was found to start at an CC internal ribosome entry site (IRES). The IRES element was found to be CC cell cycle regulated, p58 is produced during the G2/M stage of the cell CC cycle. The present sequence is the PITSLRE protein kinase IRES. The CC IRES sequence and fragments of it may be used to induce cell cycle CC dependent initiation of translation in eukaryotic cells. Vectors CC containing the IRES may be used for the preparation of compositions for CC the treatment of and/or prevention of a disease by gene therapy. Such CC diseases may be cancer and restenosis.
XX
SQ Sequence 222 BP; 77 A; 36 C; 89 G; 20 U; 0 other;

Alignment Scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

Alignment Block:
US-09-126-945B-2_COPY_229_237 x AAZ58277 ..

Align Seq 1/1 to AAZ58277 from 1 to 222

1 SerThrSerGluGluSer 6
||||||||||||||||||
112 AGCACCAGTGAAGAATCA 129

Seq_name: /cgi2_1/ptodata/1/pubseq/tmp/pubseq_NA.tmp IAY XXXXXX

Seq_documentation_block:
ID AAA73710 standard; RNA; 222 BP.
XX
AC AAA73710;
XX
DT 07-DEC-2000 (first entry)
XX
DE RNA of human PITSLRE protein kinase gene internal ribosome entry site.
XX
KW IRES; internal ribosome entry site; PITSLRE; human; gene therapy;
KW cancer; restenosis; p58; p110; protein kinase ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT misc_signal 1..222
FT /*tag= a
FT /label= IRES
FT /note= "internal ribosome entry site"
XX
PN WO200044896-A1.
XX
PD 03-AUG-2000.
XX
PF 26-JAN-2000; 2000WO-EP000643.
XX
PR 26-JAN-1999; 99EP-0200216.
XX
PA (VLAA-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOL.
XX
PI Cornelis S, Beyaert R;
XX
DR WPI; 2000-499331/44.
XX
PT Nucleic acids encoding internal ribosome entry sequences useful for
PT directing protein expression in gene therapy procedures
XX
PS Disclosure; Page 42; 57pp; English.
XX
CC The PITSLRE protein kinase gene can be translated to give two isoforms,
CC p110 and p58. Transcription of p58 mRNA was found to start at an
CC internal ribosome entry site (IRES). The IRES element was found to be
CC cell cycle regulated, p58 is produced during the G2/M stage of the cell
CC cycle. The present sequence is the corresponding RNA of the PITSLRE
CC IRES described in AAA73710. The IRES sequence and fragments of it may be
CC used to induce cell cycle dependent initiation of translation in
CC eukaryotic cells. Vectors containing the IRES may be used for the
CC preparation of compositions for the treatment of and/or prevention of
CC a disease by gene therapy. Such diseases may be cancer and restenosis.
XX
SQ Sequence 222 BP; 77 A; 36 C; 89 G; 20 U; 0 other;

Alignment Scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

Alignment Block:
US-09-126-945B-2_COPY_229_237 x AAA73710 ..

Align Seq 1/1 to AAA73710 from 1 to 222

1 SerThrSerGluGluSer 6
||||||||||||||||||
112 AGCACCAGTGAAGAATCA 129

Seq_name: /cgi2_1/ptodata/1/pubseq/tmp/pubseq_NA.tmp IAY XXXXXX

Seq_documentation_block:
ID AAZ58277 standard; DNA; 515 BP.
XX
AC AAZ58277;
XX
DT 08-MAY-2000 (first entry)
XX
DE Rice resistance gene homologue clone R5-1.
XX
KW Rice; resistance gene homologue; RGH; transgenic plant;
KW disease resistance; crop protection; ds.
XX
OS Oryza sativa.
XX
FH Key Location/Qualifiers
FT CDS 3..515
FT /*tag= a
FT /partial
XX
PN WO200061755-A2.
XX
PD 27-JAN-2000.
XX
PF 16-JUL-1999; 99WO-US16168.
XX
PR 17-JUL-1998; 98US-0093290.
PR 15-JUL-1999; 99US-0354453.
XX
PA (PURD) PURDUE RES FOUND.
XX
PI Bennetzen JL;
XX
DR WPI; 2000-182422/16.
DR P-PSDB; AAY58845.
XX
PT New nucleic acid molecule encoding a resistance gene homologue, useful
PT for producing disease resistant transgenic plant
XX
PS Claim 1(a); Page 64-65; 98pp; English.
XX
CC The present sequence is that of R5-1, a non full length clone of a
CC novel rice resistance gene homologue (RGH). R5-1 was obtained
CC from rice genomic DNA by PCR amplification. The invention
CC provides 14 novel maize, sorghum and rice RGH gene families (see
CC AAZ58278-95) and their protein products (see AAY58842-54). Partial RGH
CC sequences isolated from rice were designated R5-1, R5-2, R5-3 and
CC R5-4. The novel RGH sequences can be used for the production of

```
CC   transgenic plants (and cells, tissues and seeds) having improved
CC   resistance to disease.  Expression of the novel RGR gene sequences
CC   confers disease resistance to a plant by interacting with the
CC   complementing phytopathogen avirulence gene product released into
CC   the plant by the invading plant pathogen.  The methods of the
CC   invention find use in controlling plant pests, including fungal
CC   pathogens, viruses, nematodes and insects.
XX
SQ   Sequence 515 BP; 152 A; 80 C; 143 G; 150 T; 0 other;


alignment scores:
                Quality:     6.00              Length:       6
                  Ratio:    1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2 (COPY 229-237) x AAZ58277

 Align seq 1/1  to: AAZ58277  from: 1  to: 515

        2 ThrSerGluGluSerTrp 7
          ||||||||||||||||||
      387 ACTTCTGAAGAATCATGG 404

seq name: /tmp5/pipda/pes-seq-3p-ss-pg_NA.  0 bA: AAE40290

seq documentation block:
ID   AAE40290 standard; DNA; 550 BP.
XX
AC   AAE40290;
XX
DT   30-APR-2001  (first entry)
XX
DE   ICP27 modified promoter active sequence and expression construct.
XX
KW   Gal4; transcription factor; HSV1; vector; ICP27; promoter;
KW   gene therapy; ss.
XX
OS   Chimeric - Herpes simplex virus type 1.
OS   Chimeric - synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..18
FT                   /*tag=  a
FT                   /partial
FT                   /note= "3' end of UL54 gene coding region"
FT   3'UTR           19..135
FT                   /*tag=  b
FT                   /note= "UL54 gene 3' untranslated region"
FT   protein_bind    185..260
FT                   /*tag=  c
FT                   /bound_moiety= "GAL4"
FT   promoter        261..547
FT                   /*tag=  d
FT                   /note= "modified ICP27 promoter"
FT   CDS             536..550
FT                   /*tag=  d
FT                   /partial
FT                   /note= "5' end of ICP27 gene coding sequence; the
FT                         expression construct for ICP27 can continue
FT                         with the remainder of the coding sequence
FT                         of ICP27 (omitted in the specification for
FT                         brevity)"
XX
PN   WO200105992-A1.
XX
PD   25-JAN-2001.
XX
PF   17-JUL-2000; 2000WO-GB02739.
XX
PR   16-JUL-1999;  99GB-0016702.
XX
```

```
PA   (PANI ) PANIAB PHARM KES LTD.
XX
PI   Howard MK;
XX
DR   WPI; 2001-156913/16.
XX
PT   Novel mutant virus comprising modified gene promoter that is
PT   non-responsive to normal virus and host transcription factor but
PT   responsive to heterologous transcription factor, operably linked to
PT   essential viral gene
XX
PS   Disclosure; Appendix A.4.1; 24pp; English.
XX
CC   The present sequence is that of a herpes simplex virus type 1
CC   (HSV1) modified ICP27 active promoter and expression construct.
CC   The modified promoter is obtained by deleting a TAATGARAT element
CC   of the native promoter, and replacing it with 1-5 copies of a Gal4
CC   binding sequence (see AAF30282-83).  The invention relates to a novel
CC   mutant virus, e.g. HSV1 or HSV2, that comprises a modified gene
CC   promoter responsive to a Gal4 containing transcription factor,
CC   operably linked to an essential viral gene, e.g. an immediate early
CC   gene.  The modified promoter is non-responsive to normal virus and
CC   host transcription factors but is responsive to a heterologous
CC   transcription factor, e.g. Gal4 or a Gal4-VP16 fusion, so that the
CC   mutant virus is genetically disabled in the absence of the
CC   transcription factor.  The virus can be used as a vector for
CC   delivery of a heterologous gene to a target cell, e.g. for delivery
CC   of a gene to the central nervous system (CNS) of a subject in vivo
CC   (claimed).  The virus vectors are essentially non-toxic and are
CC   producible in good yield in complementing cell lines.  The
CC   mutations prevent expression and consequential cytotoxicity of the
CC   immediate early genes if the vector has become latent within the
CC   CNS of a treated subject.  The vectors also reduce any risk of
CC   reactivating any latent wild type virus that may already be
CC   present in the CNS.
XX
SQ   Sequence 612 BP; 134 A; 121 C; 165 G; 111 T; 0 other;


alignment scores:
                Quality:     6.00              Length:       6
                  Ratio:    1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_229_237 x AAE40290_1:1 ..

 Align seq 1/1  to reverse of: AAE40290  from: 1  to: 550

        1 SerThrSerGluGluSer 6
          ||||||||||||||||||
      265 AGTTACTTCGGAGGAGAGT 248

seq_name: /tmp5/pipda/pes-seq-3p-ss-pg_NA.rna..dAT.AAV04607

seq documentation block:
ID   AAV04607 standard; cDNA; 612 BP.
XX
AC   AAV04607;
XX
DT   02-JUL-1998  (first entry)
XX
DE   Flea serine protease gene SEQ ID NO:146.
XX
KW   Flea; Siphonaptera; protease; vaccine; prophylaxis; infestation;
KW   immunoglobulin protease; larvae; host animal; ss.
XX
OS   Siphonaptera.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..612
FT                   /*tag=  a
FT                   /product= "Serine protease"
```

```
PT                    Xxxxx "xxx xxxxxxxxx xx xxx"
XX
PN   WO9740058-A1.
XX
PD   30-OCT-1997.
XX
PF   24-APR-1997;    97WO-US06121.
XX
PR   04-APR-1997;    97US-0042945.
PR   24-APR-1996;    96US-0639075.
PR   15-NOV-1996;    96US-0749699.
XX
PA   (HESK-) HESKA CORP.
XX
PI   Frank GP,  Gaines PJ,  Grieve RB,  Hunter SW,  Rushlow KE;
PI   Silver G,  Stiegler GL;
XX
DR   WPI; 1998-076762/07.
DR   P-PSDB; AAW41982.
XX
PT   New flea protease genes and proteins - used in vaccine compositions
PT   for the prophylaxis and treatment of flea infestation, especially in
PT   cats or dogs
XX
PS   Claim 1; Page 272-273; 418pp; English.
XX
CC   The present sequence encodes a novel flea serine protease. The
CC   protease, its mimetopes, antibodies (Ab) and inhibitors of the
CC   protein, as well as the DNA encoding the protein, may all be used in
CC   therapeutic compositions to reduce flea protease activity (especially
CC   immunoglobulin protease) and so reduce flea infestation, especially in
CC   cats or dogs. Alternatively, flea larvae may ingest the faeces of adult
CC   fleas which comprises anti-protease antibodies produced by a host animal
CC   in response to administration of the protein. Therapeutic compositions
CC   may further comprise a compound that reduces haematophagous ectoparasite
CC   burden by a method other than by reducing flea immunoglobulin protease
CC   activity. The novel flea DNA encoding the protein can also used to
CC   produce recombinant protein, and fragments of it are used as probes and
CC   primers for identification and isolation of related sequences, also as
CC   antisense, triplex-forming agents and ribozymes for inhibition of the
CC   synthesis of the protein. Ab are also useful for screening expression
CC   libraries, to purify the protein and to target cytotoxins to fleas.
XX
SQ   Sequence 612 BP; 211 A; 92 C; 128 G; 181 T; 0 other;


alignment_scores:
            Quality:    6.00              Length:      6
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_229_237 x AAV04607   ..

 Align seq 1/1  to: AAV04607  from: 1  to: 612

       1 SerThrSerGluGluSer 6
         ||||||||||||||||||
      52 TCTACTTCAGAAGAATCA 69

seq_name: /cds1/readata/geneseq_pro_seqs/NA2003.DAT AAV04607

seq_documentation_block:
ID   AAC90888 standard; cDNA; 612 BP.
XX
AC   AAC90888;
XX
DT   19-MAR-2001  (first entry)
XX
DE   Flea serine protease nISP29-612 nucleotide sequence #146.
XX
KW   Flea; cat flea; serine protease; aminopeptidase; cysteine protease;
KW   flea infestation; proteolytic; insecticide; vaccine; cat; dog; dog flea;
KW   Ctenocephalides felis; Ctenocephalides canis; Pulex irritans;
KW   ectoparasite; SS.
XX
OS   Ctenocephalides felis.
XX
PN   US6150125-A.
XX
PD   21-NOV-2000.
XX
PF   24-APR-1996;    96US-0639075.
XX
PR   13-DEC-1994;    94US-0806482.
PR   18-OCT-1994;    94US-0326773.
PR   07-JUN-1995;    95US-0482130.
PR   07-JUN-1995;    95US-0484211.
PR   07-JUN-1995;    95US-0485443.
PR   07-JUN-1995;    95US-0485455.
PR   15-APR-1997;    97WO-US14442.
XX
PA   (HESK-) HESKA CORP.
XX
PI   Gaines PJ,  Silver G,  Rushlow KE,  Hunter SW,  Frank GR,  Stiegler GL;
PI   Grieve RB;
XX
DR   WPI; 2001-136424/14.
DR   P-PSDB; AAB50626.
XX
PT   New isolated flea proteins with proteolytic activity, useful for
PT   preventing and reducing flea infestations in mammals especially cats
PT   and dogs -
XX
PS   Claim 4; Column 225-228; 150pp; English.
XX
CC   The present invention describes isolated flea serine protease,
CC   aminopeptidase and cysteine protease proteins (I). Also described is a
CC   method for identifying a compound (II) capable of inhibiting flea
CC   protease activity comprising: (a) contacting (I) with a protease
CC   substrate and a putative inhibitory compound, where (I) has proteolytic
CC   activity in the absence of the compound; and (b) determining if the
CC   compound inhibits protease activity by detecting cleavage of the
CC   protease substrates; where decreased cleavage of the protease substrate
CC   indicates an inhibitory compound. (I), nucleic acid molecules encoding
CC   (I), and antibodies immunospecific for (I) and (II) are useful for
CC   preventing and reducing flea infestations, particularly the species
CC   Ctenocephalides felis (cat fleas), Ctenocephalides canis (dog fleas) and
CC   Pulex irritans, in animals, preferably cats and dogs. They are also
CC   useful for reducing infestation by other ectoparasites, preferably
CC   mosquitoes, midges, sandflies, blackflies, ticks and rhodnius. AAC90818
CC   to AAC90913 and AAB50551 to AAB50644 represent sequences used in the
CC   exemplification of the present invention.
XX
SQ   Sequence 612 BP; 211 A; 92 C; 128 G; 181 T; 0 other;


alignment_scores:
            Quality:    6.00              Length:      6
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_229_237 x AAC90888   ..

 Align seq 1/1  to: AAC90888  from: 1  to: 612

       1 SerThrSerGluGluSer 6
         ||||||||||||||||||
      52 TCTACTTCAGAAGAATCA 69

seq_name: /cds1/readata/geneseq_pro_seqs/NA2003.DAT AAA73714

seq_documentation_block:
ID   AAA73714 standard; DNA; 660 BP.
XX
```

```
AC   AAA73714;
XX
DT   07-DEC-2000  (first entry)
XX
DE   Human PITSLRE protein kinase gene IRES and upstream sequence.
XX
KW   IRES; internal ribosome entry site; PITSLRE; human; gene therapy;
KW   cancer; restenosis; p58; p110; protein kinase; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   misc_signal     439..660
FT                   /*tag= a
FT                   /label= IRES
FT                   /note= "internal ribosome entry site"
XX
PN   WO200044896-A1.
XX
PD   03-AUG-2000.
XX
PF   26-JAN-2000; 2000WO-EP000643.
XX
PR   26-JAN-1999;  99EP-0200216.
XX
PA   (VLAA ) VLAAMS INTERUNIVERSITAIR INST BIOTECHN&C.
XX
PI   Cornelis S, Beyaert R;
XX
DR   WPI; 2000-499441/44.
XX
PT   Nucleic acids encoding internal ribosome entry sequences useful for
PT   directing protein expression in gene therapy procedures
XX
PS   Claim 10; Page 84; 57pp; English.
XX
CC   The PITSLRE protein kinase gene can be translated to give two isoforms,
CC   p110 and p58. Transcription of p58 mRNA was found to start at an
CC   internal ribosome entry site (IRES). The IRES element was found to be
CC   cell cycle regulated, p58 is produced during the G2/M stage of the cell
CC   cycle. The present sequence contains the PITSLRE protein kinase
CC   IRES and upstream nucleotides. The IRES sequence and fragments of it
CC   may be used to induce cell cycle dependent initiation of translation in
CC   eukaryotic cells. Vectors containing the IRES may be used for the
CC   preparation of compositions for the treatment of and/or prevention of
CC   a disease by gene therapy. Such diseases may be cancer and restenosis.
XX
SQ   Sequence 660 BP; 210 A; 128 C; 261 G; 61 T; 0 other;


Alignment scores:
              Quality:    6.00              Length:       6
                Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-126-945B-2 (COPY_229_237) x AAA73714   ..

 Align seq 1/1   to: AAA73714  from: 1  to: 660

        1 SerThrSerGluGluSer 6
          ||||||||||||||||||
      550 AGCACCAGTGAAGAATCA 567

Seq name: /cgi6/ptodata/1/pubseq/pubseq_2/AA/AAA73714.pep.ab1.AAA73714

Seq documentation block:
ID   AAC41199 standard; DNA; 708 BP.
XX
AC   AAC41199;
XX
DT   17-OCT-2000  (first entry)
XX
```

```
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 31013.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134470.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135453.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
```

PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.

PR 30-AUG-1999; 99US-0151308.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152364.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158869.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_229_237 x AAC41199 ..

Align seq 1/1 to: AAC41199 from: 1 to: 708

1 SerThrSerGluGluSer 6
||||||||||||||||||
428 AGTACAAGCGAGGAAAGC 445

seq_header: /cgi/jt/align/seq -g/tmp/ge_2NA2000_DAT_AAF20947

seq_documentation_block:
ID AAF20947 standard; DNA; 720 BP.
XX
AC AAF20947;
XX
DT 14-MAR-2001 (first entry)
XX
DE Human interleukin 1 polynucleotide treatment #2514.

XX
KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200062736-A2.
XX
PD 26-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-US08020.
XX
PR 06-APR-1999; 99US-0127958.
XX
PA (UYEC-) UNIV EAST CAROLINA.
PA (NYCE/) NYCE J W.
XX
PI Nyce JW;
XX
DR WPI; 2000-679549/66.
XX
PT Low adenosine (A) content antisense oligonucleotides which do not
PT trigger adenosine receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions -
XX
PS Disclosure; Page 233; 1592pp; English.
XX
CC The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down regulate the
CC expression and or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18444 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.
XX
SQ Sequence 720 BP; 242 A; 165 C; 165 G; 148 T; 0 other;

alignment scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2-COPY-229-237 x AAF20947 ..

Align set 1/1 to: AAF20947 from: 1 to: 720

1 SerIleSerGlnGlnSer 6
|||| ||||||||||||
4co 1CIACAICIGAAGAAICA 486

OM nt: US-09-126-945B-2_COPY_229_237 [illegible] Issued_Patents_NA:* [illegible]

Date: [illegible] 2001 [illegible] PM

Above results were produced by the GenCore software [illegible]
Copyright [illegible]

Command line parameters:
-MODEL frame+_p2n.model -DEV=xlp
-Q [illegible] /app_query.fasta 1.1267
-DB Issued_Patents_NA -QFMT fastap -SUFFIX olip2n.rni
-GAPOP 4.500 -GAPEXT 0.050 -MINMATCH 0.100 -LOOPCL 0.000
-LOOPEXT 0.000 -QGAPOP 4.500 -QGAPEXT 0.050 -XGAPOP 60.000
-XGAPEXT 60.000 -FGAPOP 6.000 -FGAPEXT 7.000 -YGAPOP 60.000
-YGAPEXT 60.000 -DELOP 6.000 -DELEXT 7.000 -START 1
-MATRIX oligo -TRANS human40.cdi -LIST 45 -DOCALIGN 200
-THR_SCORE quality -THR_MIN 1 -ALIGN 15 -MODE LOCAL -OUTFMT pto
-NORM ext -MINLEN 0 -MAXLEN 2000000000
-USER US09126945_@CGN1_1_91 -NCPU 6 [illegible] No_XGAPEXT
-WAIT -THREADS 1

Search information block:
query: US-09-126-945B-2_COPY_229_237
query length: 9
Database: Issued_Patents_NA:*
Database sequences: 324599
Database length: 94655562
Search time (sec): 322.720000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000

score_list:
Sequence Strd Orig ZScore EScore Len ! Documentation ...
[illegible]6A_COMB seq:US-08-906-769-146 + 6.00 120.43 43.07 612 !
[illegible]6A_COMB seq:US-08-906-616-146 + 6.00 120.43 43.07 612 !
[illegible]6A_COMB seq:US-08-[illegible]-146 + 6.00 120.43 43.07 612 !
[illegible]
[illegible]A_COMB seq:US-08-[illegible]-12 + 6.00 106.45 258.76 251 !
[illegible]
[illegible]A_COMB seq:US-08-428-256-1 + 6.00 104.00 354.37 360 !
[illegible]
[illegible]B_COMB seq:US-09-060-756-653 + 6.00 102.81 412.92 429 !
[illegible]
[illegible]A_COMB seq:US-08-680-726A-3 6.00 101.10 514.43 552 !
[illegible]
[illegible]5A_COMB seq:US-08-019-904-2 + 6.00 100.43 560.45 609 !
[illegible]
[illegible]PCTUS_COMB seq:PCT-US94-00928-1 + 6.00 100.43 560.45 609
[illegible]
[illegible]_COMB seq:US-09-061-304-2 + 6.00 100.01 591.62 648 !
[illegible]
[illegible]5B_COMB seq:US-08-489-666[illegible] + 6.00 99.[illegible] 594.01 651

[illegible]6A_COMB seq:US-08-906-769-146

seq_documentation_block:
; Sequence 146, Application US/08906769
; Patent No. 6077687
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Silver, Gary
; TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
; TITLE OF INVENTION: MOLECULES AND USES THEREOF
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/906,769
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: [illegible] 08/639,075
; FILING DATE: 24-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: [illegible] Gary J.
; REGISTRATION NUMBER: 32,920
; REFERENCE/DOCKET NUMBER: 2618-28-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 146:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 612 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..612
US-08-906-769-146

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_229_237 x US-08-906-769-146 ..

Align seq 1/1 to: US-08-906-769-146 from: 1 to: 612

```
        1 SerThrSerGluGluSer 6
          ||||||||||||||||||
       52 TCTACTTCAGAAGAATCA 69

seq_name: /cgn2_6/ptmp/data/2/ina/A_C_MB_seq_US_08_906_616_146

seq_documentation_block:
; Sequence 146, Application US/08906616
; Patent No. 6121045
;  GENERAL INFORMATION:
;     APPLICANT:  Grieve, Robert B.
;     APPLICANT:  Rushlow, Keith E.
;     APPLICANT:  Wu Hunter, Shirley
;     APPLICANT:  Frank, Glenn R.
;     APPLICANT:  Stiegler, Gary
;     APPLICANT:  Gaines, Patrick J.
;     APPLICANT:  Silver, Gary
;     TITLE OF INVENTION:  FLEA AMINOPEPTIDASE PROTEINS AND USES THEREOF
;     NUMBER OF SEQUENCES:  190
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Sheridan Ross P.C.
;       STREET:  1700 Lincoln Street, Suite 3500
;       CITY:  Denver
;       STATE:  Colorado
;       COUNTRY:  USA
;       ZIP:  80203
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/906,616
;       FILING DATE:  05-AUG-1997
;       CLASSIFICATION:  536
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Connell, Gary J.
;       REGISTRATION NUMBER:  32,020
;       REFERENCE/DOCKET NUMBER:  2618-25 C2 3
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (303) 863-9700
;       TELEFAX:  (303) 863-0223
;   INFORMATION FOR SEQ ID NO:  146:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  612 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..612
US 08 906 616-146


alignment_scores:
             Quality:    6.00              Length:      6
               Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_229_237 x US-08-906-616-146    ..

 Align seq 1/1  to: US-08-906-616-146  from: 1  to: 612

        1 SerThrSerGluGluSer 6
          ||||||||||||||||||
       52 TCTACTTCAGAAGAATCA 69

Seq_name: /cgn2_6/ptmp/data/2/ina/A_C_MB... ;US-08-906-616-146

seq_documentation_block:
; Sequence 146, Application US/08639075A
; Patent No. 6150125
;  GENERAL INFORMATION:
;     APPLICANT:  Grieve, Robert B.
;     APPLICANT:  Rushlow, Keith E.
;     APPLICANT:  Wu Hunter, Shirley
;     APPLICANT:  Frank, Glenn R.
;     APPLICANT:  Stiegler, Gary
;     APPLICANT:  Gaines, Patrick J.
;     APPLICANT:  Silver, Gary
;     TITLE OF INVENTION:  FLEA PROTEASE PROTEINS, NUCLEIC ACID
;     TITLE OF INVENTION:  MOLECULES AND USES THEREOF
;     NUMBER OF SEQUENCES:  190
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Sheridan Ross & McIntosh
;       STREET:  1700 Lincoln Street, Suite 4500
;       CITY:  Denver
;       STATE:  Colorado
;       COUNTRY:  USA
;       ZIP:  80203
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/639,075A
;       FILING DATE:  24-APR-1996
;       CLASSIFICATION:  424
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Connell, Gary J.
;       REGISTRATION NUMBER:  32,020
;       REFERENCE/DOCKET NUMBER:  2618-25-C2
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (303) 863-9700
;       TELEFAX:  (303) 863-0223
;   INFORMATION FOR SEQ ID NO:  146:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  612 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..612
US 08 639 075A 146


alignment_scores:
             Quality:    6.00              Length:      6
               Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_229_237 x US-08-639-075A-146    ..

 Align seq 1/1  to: US 08 639 075A 146  from: 1  to: 612

        1 SerThrSerGluGluSer 6
          ||||||||||||||||||
       52 TCTACTTCAGAAGAATCA 69

Seq_name: /cgn2_6/ptmp/data/2/ina/60/COMB.seq.US_09_012_431_146

seq_documentation_block:
; Sequence 146, Application US/09012431
; Patent No. 6180383
;  GENERAL INFORMATION:
;       APPLICANT: Grieve, Robert B.
;                  Rushlow, Keith E.
;                  Wu Hunter, Shirley
;                  Frank, Glenn R.
;                  Stiegler, Gary
```

```
;                    Gaines, Patrick J.
;                    Silver, Gary
;       TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
;                           MOLECULES AND USES THEREOF
;       NUMBER OF SEQUENCES: 190
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Sheridan Ross & McIntosh
;            STREET: 1700 Lincoln Street, Suite 3500
;            CITY: Denver
;            STATE: Colorado
;            COUNTRY: USA
;            ZIP: 80203
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/012,692
;            FILING DATE: 23-Jan-1998
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US 08/639,075
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Connell, Gary J.
;            REGISTRATION NUMBER: 32,020
;            REFERENCE/DOCKET NUMBER: 2618-25-C2
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (303) 863-9700
;            TELEFAX: (303) 863-0223
;   INFORMATION FOR SEQ ID NO: 146:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 612 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: cDNA
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  1..612
;       SEQUENCE DESCRIPTION: SEQ ID NO: 146:
US-09-012-431-146


alignment_scores:
            Quality:    6.00              Length:      6
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_229_237 x US-09-012-431-146    ..

 Align seq 1/1  to: US-09-012-431-146  from: 1  to: 612

       1 SerThrSerGluGluSer 6
         ||||||||||||||||||
      52 TCTACTTCAGAAGAATCA 69

seq_name: Query Sequence 1/2/final/P_COMB seq:US-09-012-692-146

seq_documentation_block:
; Sequence 146, Application US/09012692
; Patent No. 6214579
;  GENERAL INFORMATION:
;    APPLICANT:  Grieve, Robert B.
;    APPLICANT:  Rushlow, Keith E.
;    APPLICANT:  Wu Hunter, Shirley
;    APPLICANT:  Frank, Glenn R.
;    APPLICANT:  Stiegler, Gary
;    APPLICANT:  Gaines, Patrick J.
;    APPLICANT:  Silver, Gary
;    TITLE OF INVENTION:  FLEA PROTEASE PROTEINS, NUCLEIC ACID
;    TITLE OF INVENTION:  MOLECULES AND USES THEREOF
;    NUMBER OF SEQUENCES:  190
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheridan Ross & McIntosh
;      STREET:  1700 Lincoln Street, Suite 3500
;      CITY:  Denver
;      STATE:  Colorado
;      COUNTRY:  USA
;      ZIP:  80203
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/012,692
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/639,075
;      FILING DATE:  24-APR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Connell, Gary J.
;      REGISTRATION NUMBER:  32,020
;      REFERENCE/DOCKET NUMBER:  2618-25-C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 863-9700
;      TELEFAX:  (303) 863-0223
;  INFORMATION FOR SEQ ID NO:  146:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  612 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..612
US-09-012-692-146


alignment_scores:
            Quality:    6.00              Length:      6
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_229_237 x US-09-012-692-146    ..

 Align seq 1/1  to: US-09-012-692-146  from: 1  to: 612

       1 SerThrSerGluGluSer 6
         ||||||||||||||||||
      52 TCTACTTCAGAAGAATCA 69

seq_name: Query Sequence 1/2/final/P_COMB seq:US-08-906-613-146

seq_documentation_block:
; Sequence 146, Application US/08906613
; Patent No. 5792836
;  GENERAL INFORMATION:
;    APPLICANT:  Grieve, Robert B.
;    APPLICANT:  Rushlow, Keith E.
;    APPLICANT:  Wu Hunter, Shirley
;    APPLICANT:  Frank, Glenn R.
;    APPLICANT:  Stiegler, Gary
;    APPLICANT:  Gaines, Patrick J.
;    APPLICANT:  Silver, Gary
;    TITLE OF INVENTION:  FLEA PROTEASE PROTEINS, NUCLEIC ACID
;    TITLE OF INVENTION:  MOLECULES AND USES THEREOF
;    NUMBER OF SEQUENCES:  190
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheridan Ross & McIntosh
```

```
;       STREET:  1700 Lincoln Street, Suite 3500
;       CITY:  Denver
;       STATE:  Colorado
;       COUNTRY:  USA
;       ZIP:  80203
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/[illegible]
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/639,025
;       FILING DATE:  24 APR 1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Connell, Gary J.
;       REGISTRATION NUMBER:  32,020
;       REFERENCE/DOCKET NUMBER:  2618.25-CA
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (303) 863-9700
;       TELEFAX:  (303) 863-0223
;   INFORMATION FOR SEQ ID NO:  146:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  612 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..612
US 08 906 613 146


alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US 09 126 945B 2 copy 229_237 x US 08 906 613 146  ..

 Align seq 1/1  to reverse of: US 08 906 613 146  from: 1  to: 612

        1 SerThrSerGluGluSer 6
          ||||||||||||||||||
      52 TCTACTTCAGAAGAATCA 69

seq name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US 08 805 118 2

seq documentation block:
; Sequence 2, Application US/08805118
; Patent No. 5985604
;  GENERAL INFORMATION:
;     APPLICANT:  Lal, Preeti
;     APPLICANT:  Bandman, Olga
;     TITLE OF INVENTION:  NOVEL HUMAN [illegible]-DEPENDENT
;     TITLE OF INVENTION:  PHOSPHATE CO-TRANSPORTER
;     NUMBER OF SEQUENCES:  7
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;       STREET:  3174 Porter Drive
;       CITY:  Palo Alto
;       STATE:  CA
;       COUNTRY:  US
;       ZIP:  94304
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/805,118
;       FILING DATE:  Filed Herewith
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Billings, Lucy J.
;       REGISTRATION NUMBER:  36,749
;       REFERENCE/DOCKET NUMBER:  PF-0221 US
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415 855-0555
;       TELEFAX:  415 845-4166
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1643 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     IMMEDIATE SOURCE:
;       LIBRARY:  BRAITUT02
;       CLONE:  754412
US 08 805 118 2


alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US 09 126 945B 2_copy_229_237 x US 08 805 118 2  ..

 Align seq 1/1  to reverse of: US 08 805 118 2  from: 1  to: 1643

        1 SerThrSerGluGluSer 6
          ||||||||||||||||||
      425 AGCACTTCAGAGGAATCA 408

seq name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US 08 724 494A 19

seq documentation block:
; Sequence 19, Application US/08724494A
; Patent No. 5872237
;  GENERAL INFORMATION:
;     APPLICANT:  Feder, John N.
;     APPLICANT:  Kronmal, Gregory S.
;     APPLICANT:  Lauer, Peter M.
;     APPLICANT:  Ruddy, David A.
;     APPLICANT:  Thomas, Winston
;     APPLICANT:  Tsuchihashi, Zenta
;     APPLICANT:  Wolff, Roger K.
;     TITLE OF INVENTION:  Mammalian Transcript Map:  No. 5872237-01
;     TITLE OF INVENTION:  Sequences and Antibodies Thereto
;     NUMBER OF SEQUENCES:  41
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  TOWNSEND and TOWNSEND and CREW LLP
;       STREET:  Two Embarcadero Center, 8th Floor
;       CITY:  San Francisco
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  94111-3834
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.40
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/724,494A
;       FILING DATE:  01-OCT-1996
;       CLASSIFICATION:  536
```

```
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Fitts, Renee A.
;       REGISTRATION NUMBER:  35,136
;       REFERENCE/DOCKET NUMBER:  017957-000100
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415-576-0200
;       TELEFAX:  415-576-0300
;   INFORMATION FOR SEQ ID NO:  19:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1780 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  not relevant
;       TOPOLOGY:  not relevant
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  1..1780
;       OTHER INFORMATION:  /note= "cDNA 228"
US-08-724-394A-19


Alignment_scores:
            Quality:     6.00              Length:       6
              Ratio:  1.000                 Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment_block:
 US-09-126-945B-2_COPY_229_237 x US-08-724-394A-19/rev   ..

 Align seq 1/1  to reverse of: US-08-724-394A-19  from: 1  to: 1780

       1 SerThrSerGluGluSer 6
         ||| ||||||||||||||
     567 AGCACTCACAGGAATCA 550

Seq_name: /home/biomed/private/2_Jan/FA_2_MB.seq;US-09-126-945b-2 237

seq_documentation_block:
; Sequence 1, Application US/08924137
; Patent No. 6083716
;  GENERAL INFORMATION:
;     APPLICANT:  Wilson, James M.
;     APPLICANT:  Farina, Steven F.
;     APPLICANT:  Fisher, Krishna J.
;     TITLE OF INVENTION:  Chimpanzee Adenovirus Vectors
;     NUMBER OF SEQUENCES:  5
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Howson and Howson
;       STREET:  Spring House Corporate Cntr., P.O. Box 457
;       CITY:  Spring House
;       STATE:  Pennsylvania
;       COUNTRY:  United States of America
;       ZIP:  19477
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/924,137
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 60/024,700
;       FILING DATE:  06-SEP-1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Bak, Mary E.
;       REGISTRATION NUMBER:  31,215
;       REFERENCE/DOCKET NUMBER:  CNVPN.021P1E1P5A
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  215-540-9200
;       TELEFAX:  215-540-5818
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  35524 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  cDNA
US-08-923-137-1


Alignment_scores:
            Quality:     6.00              Length:       6
              Ratio:  1.000                 Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment_block:
 US-09-126-945B-2_COPY_229_237 x US-08-923-137-1   ..

 Align seq 1/1  to: US-08-923-137-1  from: 1  to: 35524

       4 GluGluSerTrpThrAsp 9
         |||||||||||||||||
   19158 GAGGAAGCTTGGACTGAT 19175

Seq_name: /home/biomed/private/2_Jan/FA_2_MB.seq;US-09-513-729B-37

seq_documentation_block:
; Sequence 37, Application US/09513729B
; Patent No. 6165791
; GENERAL INFORMATION:
;   APPLICANT:  Ian Popoff
;   APPLICANT:  Jacqueline Wyatt
;   TITLE OF INVENTION:  ANTISENSE MODULATION OF E2F TRANSCRIPTION FACTOR 3 EXPRES
;   FILE REFERENCE:  RIS-0112
;   CURRENT APPLICATION NUMBER:  US/09/513,729B
;   CURRENT FILING DATE:  2000-02-24
;   NUMBER OF SEQ ID NOS:  88
; SEQ ID NO 37
;   LENGTH: 20
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Antisense oligonucleotide
US-09-513-729B-37


Alignment_scores:
            Quality:     5.00              Length:       5
              Ratio:  1.000                 Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment_block:
 US-09-126-945B-2_COPY_229_237 x US-09-513-729B-37/rev   ..

 Align seq 1/1  to reverse of: US-09-513-729B-37  from: 1  to: 20

       2 ThrSerGluGluSer 6
         ||||||||||||||
      18 ACTTCAGAAGAAAGC 4

Seq_name: /home/biomed/private/2_Jan/FA_2_MB.seq;US-09-609-913-200

seq_documentation_block:
; Sequence 200, Application US/09009913
; Patent No. 6087485
;  GENERAL INFORMATION:
;     APPLICANT:  Axys Pharmaceuticals, Inc.
;     TITLE OF INVENTION:  Asthma Related Genes
;     NUMBER OF SEQUENCES:  439
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Bozicevic & Reed, LLP
;       STREET:  285 Hamilton Ave, Suite 200
;       CITY:  Palo Alto
;       STATE:  CA
```

```
;       COUNTRY:  USA
;       ZIP:  94401
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/009,913
;       FILING DATE:  21-JAN-1998
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Sherwood, Pamela J
;       REGISTRATION NUMBER:  36,677
;       REFERENCE/DOCKET NUMBER:  SEQ-4P
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  650-327-3241
;       TELEFAX:  650-327-3241
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  200:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  22 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-09-009-913-200


alignment scores:
              Quality:     5.00              Length:       5
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945b-2_COPY_229_237 x US-09-009-913-200  ..

 Align seq 1/1  to: US-09-009-913-200  from: 1  to: 22

        1 SerThrSerGluGlu 5
          |||||||||||||||
        1 AGCACTTCTCAGGAA 15

Seq_name: "US-09-126-945b-2_COPY_229_237" Ref:"US-09-009-913-200" MM:swgp; 1 / 22 #1 *

seq_documentation_block:
; Sequence 30, Application US/07989847
; Patent No. 5866364
;   GENERAL INFORMATION:
;     APPLICANT:  Israel, David
;     APPLICANT:  Wolfman, Neil M.
;     TITLE OF INVENTION:  Recombinant Bone Morphogenetic Protein
;     TITLE OF INVENTION:  Heterodimers, Compositions and Methods of Use.
;     NUMBER OF SEQUENCES:  40
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Legal Affairs, Genetics Institute, Inc.
;       STREET:  87 CambridgePark Drive
;       CITY:  Cambridge
;       STATE:  MA
;       COUNTRY:  USA
;       ZIP:  02140-2387
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Tape
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/989,847
;       FILING DATE:
;       CLASSIFICATION:  436
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Kapinos, Ellen J.
;       REGISTRATION NUMBER:  32,245
;       REFERENCE/DOCKET NUMBER:  GI 5192B
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617 876 1170
;       TELEFAX:  617 876 5851
;   INFORMATION FOR SEQ ID NO:  30:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  27 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-07-989-847-30


alignment scores:
              Quality:     5.00              Length:       5
                Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945b-2_COPY_229_237 x US-07-989-847-30  ..

 Align seq 1/1  to: US-07-989-847-30  from: 1  to: 27

        1 SerThrSerGluGlu 5
          |||||||||||||||
        5 TGAACCTCGGAGGAG 19

Seq_name: "US-09-126-945b-2_COPY_229_237" Ref:"/cgn2/ftp/GB_COMB.seq:US-08-469-411-30

seq_documentation_block:
; Sequence 30, Application US/08469411
; Patent No. 6190880
;   GENERAL INFORMATION:
;         APPLICANT: Israel, David
;                    Wolfman, Neil M.
;         TITLE OF INVENTION: Recombinant Bone Morphogenetic Protein
;                             Heterodimers, Compositions and Methods of Use.
;         NUMBER OF SEQUENCES: 40
;         CORRESPONDENCE ADDRESS:
;              ADDRESSEE: Legal Affairs, Genetics Institute, Inc.
;              STREET: 87 CambridgePark Drive
;              CITY: Cambridge
;              STATE: MA
;              COUNTRY: USA
;              ZIP: 02140-2387
;         COMPUTER READABLE FORM:
;              MEDIUM TYPE: Tape
;              COMPUTER: IBM PC compatible
;              OPERATING SYSTEM: PC DOS/MS DOS
;              SOFTWARE: PatentIn Release #1.0, Version #1.25
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/08/469,411
;              FILING DATE: 06-Jun-1995
;              CLASSIFICATION: <Unknown>
;         ATTORNEY/AGENT INFORMATION:
;              NAME: Kapinos, Ellen J.
;              REGISTRATION NUMBER: 32,245
;              REFERENCE/DOCKET NUMBER: GI 5192B-CON
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: 617-498-8622
;              TELEFAX: 617 876 5851
;   INFORMATION FOR SEQ ID NO: 30:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 27 base pairs
;              TYPE: nucleic acid
;              STRANDEDNESS: single
;              TOPOLOGY: linear
;         MOLECULE TYPE: DNA (genomic)
;         SEQUENCE DESCRIPTION: SEQ ID NO: 30:
US-08-469-411-30
```

```
alignment_scores:
            Quality:     5.00              Length:       5
              Ratio:    1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_229_237 x US-08-469-411-30  .

 Align seq 1/1  to: US-08-469-411-30  from: 1  to: 27

        1 SerThrSerGluGlu 5
          |||||||||||||||
        5 ICGACCICGGAGGAG 19

seq_name  /tmp/p[redirect.rni,PCTUS_COMB.seq PCT-US94-06456-11

seq_documentation_block:
; Sequence 11, Application PC/TUS9406456
;  GENERAL INFORMATION:
;     APPLICANT:  Beutel, Bruce A.
;     APPLICANT:  Coppola, George R.
;     APPLICANT:  Sherman, Michael I.
;     TITLE OF INVENTION:  Oligonucleotides Which Inhibit HIV Protease Function
;     NUMBER OF SEQUENCES:  60
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Carella, Byrne, Bain, Gilfillan, Cecchi, Stewart & Olstein
;       STREET:  6 Becker Farm Road
;       CITY:  Roseland
;       STATE:  New Jersey
;       COUNTRY:  USA
;       ZIP:  07068
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5 inch diskette
;       COMPUTER:  IBM PS/2
;       OPERATING SYSTEM:  PC - DOS
;       SOFTWARE:  DW4.V2
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US94/06456
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/073,873
;       FILING DATE:  09-JUN-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Olstein, Elliott M.
;       REGISTRATION NUMBER:  24,025
;       REFERENCE/DOCKET NUMBER:  23550-89
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  201-994-1700
;       TELEFAX:  201-994-1744
;   INFORMATION FOR SEQ ID NO:  11:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  98 bases
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  oligonucleotide
PCT-US94-06456-11


alignment_scores:
            Quality:     5.00              Length:       5
              Ratio:    1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_229_237 x PCT-US94-06456-11/rev  ..

 Align seq 1/1  to reverse of: PCT-US94-06456-11  from: 1  to: 98

        1 SerThrSerGluGlu 5
          |||||||||||||||
       84 TCCACTAGTGAGGAA 20

seq_name  /tmp/p[redirect.rni,PCTUS_COMB.seq PCT-US94-06456-40

seq_documentation_block:
; Sequence 40, Application PC/TUS9406456
;  GENERAL INFORMATION:
;     APPLICANT:  Beutel, Bruce A.
;     APPLICANT:  Coppola, George R.
;     APPLICANT:  Sherman, Michael I.
;     TITLE OF INVENTION:  Oligonucleotides Which Inhibit HIV Protease Function
;     NUMBER OF SEQUENCES:  60
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Carella, Byrne, Bain, Gilfillan, Cecchi, Stewart & Olstein
;       STREET:  6 Becker Farm Road
;       CITY:  Roseland
;       STATE:  New Jersey
;       COUNTRY:  USA
;       ZIP:  07068
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5 inch diskette
;       COMPUTER:  IBM PS/2
;       OPERATING SYSTEM:  PC - DOS
;       SOFTWARE:  DW4.V2
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US94/06456
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/073,873
;       FILING DATE:  09-JUN-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Olstein, Elliott M.
;       REGISTRATION NUMBER:  24,025
;       REFERENCE/DOCKET NUMBER:  23550-89
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  201-994-1700
;       TELEFAX:  201-994-1744
;   INFORMATION FOR SEQ ID NO:  40:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  98 bases
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  oligonucleotide
PCT-US94-06456-40


alignment_scores:
            Quality:     5.00              Length:       5
              Ratio:    1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_229_237 x PCT-US94-06456-40/rev  ..

 Align seq 1/1  to reverse of: PCT-US94-06456-40  from: 1  to: 98

        1 SerThrSerGluGlu 5
          |||||||||||||||
       84 TCCACTAGTGAGGAA 20
```

```
/M of: US-09-126-945B-2_COPY_229_237.fa: EST:*  -out_format . pts

Date: Sep 14, 2001  7:47 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
  -MODEL=frame_p2n.model -DEV=xlp
  -Q=/cgn2_1_0/data/seqs/US09_sp_wo_Sp_Ge_Pa/us-09-126-945b-2_copy_229_237.fa -DB=EST:* -FORMAT=fastap -SUFFIX=olip2n.rst -GAPOP 4.500
  -GAPEXT 0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT 0.000
  -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
  -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT 60.000
  -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=olig>
  -TRANS human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality
  -THR_MIN=1 -ALIGN=15  MODE LOCAL  OUTFMT pts  NORM ext  MINLEN 0
  -MAXLEN 2003000000 -USER US09126945_@CGN1_1_4645 -NCPUS=6
  -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THRPADS=1

Search information block:
Query: US-09-126-945B-2_COPY_229_237
Query length: 9
Database: EST:*
Database sequences: 10228115
Database length: 431459454
Search time (sec): 9087.630000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT=60.000


score_list:
Sequence              Strd Orig    ZScore    EScore Len  ! Documentation ..
qb_est78:BE757615      +     9.00  177.57   0.9004    539  ! BE757615 212104 MARC 2BOV Bos t
qb_est97:BG179568      +     9.00  173.51   1.52      993  ! BG179568 602328182F1 NIH_MGC_9
qb_est45:AW284602      +     7.00  141.72  89.41      327  ! AW284602 LG1_271_B09.q1_A002 Li
qb_est73:BE363525      +     7.00  140.27  107.58     406  ! BE363525 WS1_63_B02.q1_A002 Wat
qb_est50:AW679111      +     7.00  139.21  123.25     476  ! AW679111 WS1_22_G07.q1_A002 Wat
qb_qss27:AV747732      +     7.00  138.20  140.82     553  ! AV747732 IM0290B14F Mixed 10kb
qb_est76:BE597371      +     7.00  137.91  145.71     579  ! BE597371 P11_72_F10.q1_A002 Pat
qb_est89:BF574485      +     7.00  136.52  174.10     713  ! BF574485 602132022F1 NIH_MGC_81
qb_est89:BF532587      +     7.00  136.24  180.56     744  ! BF532587 602074643F1 NCI_CGAP_1
qb_est60:BF040978      -     7.00  134.24  233.27    1041  ! BF040978 CF15F03 CF Oryza sati
qb_est98:BG243672      -     7.00  134.14  236.25    1019  ! BG243672 602357762F1 NCI_CGAP_
qb_qss16:AQ752331      +     7.00  132.94  275.75    1221  ! AQ752331 BS_5565_B2_B09_97A_B1
qb_est2:AA086599       -     6.00  127.41  560.13     148  ! AA086599 mm91f05.r1 Stratagene
qb_est92:BF772068      +     6.00  126.93  595.57     159  ! BF772068 CM4-UT0041-051200-477-
qb_qss5:FP080420       -     6.00  126.93  595.57     159  ! FP080420 F.fibripes GSS sequenc
qb_est71:BE150872      -     6.00  126.85  601.93     161  ! BE150872 PM4-BT0276-160200-013-
qb_qss8:AQ200049       +     6.00  126.77  608.32     163  ! AQ200049 RPCI11 5918.TK RPCI-11
qb_est94:BF915937      -     6.00  126.41  636.92     172  ! BF915937 IL4-UT0114-081200-369-
qb_est55:BB036315      -     6.00  126.37  640.08     173  ! BB036315 RIKEN full le
qb_qss7:AQ083598       +     6.00  126.22  652.71     177  ! AQ083598 RPCI11 5607.TJ RPCI-11
qb_est4:AA286419       -     6.00  125.67  699.72     192  ! AA286419 oa29c01.r1 CharyWoodfor
qb_est21:AI550729      -     6.00  125.47  718.37     198  ! AI550729 mm91f05.y1 Stratagene
qb_est36:AV304184      -     6.00  125.17  746.19     207  ! AV304184 AV304184 RIKEN full-le
qb_est8:AA501281       +     6.00  125.05  758.50     211  ! AA501281 oh62c07.r1 Knowles Pe
qb_est94:BF923139      +     6.00  124.92  770.78     215  ! BF923139 PM3-NT0151-291100-022-
qb_est29:AT000080      -     6.00  124.86  776.69     217  ! AT000080 AT000080 Apple young f
qb_est37:AV315324      -     6.00  124.83  779.96     218  ! AV315324 AV315324 RIKEN full-le
qb_est63:BB362591      -     6.00  124.68  795.23     223  ! BB362591 BB362591 RIKEN full-le
qb_est60:BF270207      -     6.00  124.65  798.29     224  ! BF270207 BF270207 RIKEN full le
qb_est55:BB052291      -     6.00  124.59  804.37     226  ! BB052291 BB052291 RIKEN full-le
qb_est65:BB467155      +     6.00  124.47  816.52     230  ! BB467155 BB467155 RIKEN full le
qb_est105:H74727       +     6.00  124.47  816.52     230  ! H74727 272 Deletion-treated Bra
qb_est42:AW099694      +     6.00  124.38  825.62     233  ! AW099694 sd30a01.y2 sm 0012 01
qb_est41:AV025896      -     6.00  124.84  885.84     253  ! AV025896 AV025896 Mus musculus
qb_est70:BE080840      -     6.00  123.84  885.84     253  ! BE080840 QV1-BT0631-180200-078-
qb_est69:BB603191      +     6.00  123.73  897.80     257  ! BB603191 BB603191 RIKEN full-le
qb_est110:W20642       -     6.00  123.73  897.80     257  ! W20642 mb92c04.r1 Soares mouse
qb_est10:AA707175      +     6.00  123.71  900.79     258  ! AA707175 zj26e02.s1 Soares_fete
qb_est6:AA384280       -     6.00  123.63  909.72     261  ! AA384280 PS97783 Thyroid Bonn
qb_est35:AV298129      -     6.00  123.53  921.63     265  ! AV298129 AV298129 RIKEN full
qb_est3:AA160244       -     6.00  123.38  939.44     271  ! AA160244 zo82f04.r1 Stratagene
qb_est108:T00955       -     6.00  123.33  945.37     273  ! T00955 wES1o1676 Early embry
qb_est82:BF056763      -     6.00  123.30  948.33     274  ! BF056763 7k08c12.x1 NCI_CGAP
qb_est61:BB294810      -     6.00  123.26  954.24     276  ! BB294810 BB294810 RIKEN full
qb_est37:AV323240      -     6.00  123.21  960.15     278  ! AV323240 AV323240 RIKEN full

seq_name: qb_est78:BE757615

seq_documentation:
LOCUS       BE757615       539 bp    mRNA            EST       15-SEP-2000
DEFINITION  212104 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BE757615
VERSION     BE757615.1  GI:10171607
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 539)
  AUTHORS   Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
            Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
            ,W.W. and Keele,J.W.
  TITLE     Design and use of four pooled tissue normalized cDNA libraries for
            EST discovery in cattle
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.c. Vector identified by cross_match with the -minscore 18
            and  -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 94  row: G  column: 13
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .539
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 2BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from testis, thymus,
                     semitendonosus muscle, longissimus muscle, pancreas,
                     adrenal, and endometrium."
BASE COUNT       98 a     168 c     193 g      80 t
ORIGIN


alignment_scores:
            Quality:    9.00              Length:      9
              Ratio:   1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_229_237 x BE757615  ..

 Align set 1/1  to: BE757615  from: 1  to: 539

        1 SerThrSerGluGluSerTrpThrAsp 9
          |||||||||||||||||||||||||||
      447 TCGACTCAGCGAGGAGAGCTGGACGGAC 463

seq_name: qb_est97:BG179568

seq_documentation_block:
```

```
LOCUS       BG179568       991 bp    mRNA             EST       06 FEB 2001
DEFINITION  602428182T1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4429413 5',
            mRNA sequence.
ACCESSION   BG179568
VERSION     BG179568.1  GI:12686271
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 991)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10181  row: d  column: 22
            High quality sequence stop: 655.
FEATURES             Location/Qualifiers
     source          1..991
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4429413"
                     /clone_lib="NIH_MGC_91"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage resistant)"
                     /note="Organ: prostate; Vector: pCMV SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo dT primed.
                     Average insert size 1.4 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      190 a    351 c    305 g    145 t
ORIGIN


Alignment scores:
            Quality:     9.00            Length:       9
              Ratio:    1.000              Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

Alignment block:
 US-09-126-945B-2_COPY_229_237 x BG179568   ..

 Align seq 1/1  to: BG179568  from: 1  to: 991

        1 SerThrSerGluGluSerTrpThrAsp 9
          |||||||||||||||||||||||||||
      588 TCGACCAGTGAGGAGAGCTGGACCGAC 614

Seq name: gb:est45:AW284602

Seq documentation block:
LOCUS       AW284602       327 bp    mRNA             EST       19 JUL 2000
DEFINITION  LGI_271_B09_g1_A002 Light Grown 1 (LGI) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   AW284602
VERSION     AW284602.2  GI:6858094
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 327)
  AUTHORS   Cordonnier Pratt,M. M., Gingle,A., Marsala,C. and Pratt,L.H.
  TITLE     An EST database from Sorghum: light grown seedlings
  JOURNAL   Unpublished (2000)
COMMENT     On Jan 6, 2003 this sequence version replaced gi:2674446.
            Contact: Cordonnier Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: T7
            High quality sequence start: 40
            High quality sequence stop: 427
            POLYA Yes.
FEATURES             Location/Qualifiers
     source          1..327
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Light Grown 1 (LGI)"
                     /note="Organ: 10- to 14 day old light grown (greenhouse)
                     seedlings. Vector: Lambda Zap, site 1: SalI, site 2: EcoRI
                     ; The library was made from poly A RNA in the cloning
                     vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT      102 a     59 c     71 g     95 t
ORIGIN


Alignment scores:
            Quality:     7.00            Length:       7
              Ratio:    1.000              Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

Alignment Block:
 US-09-126-945B-2_COPY_229_237 x AW284602   ..

 Align seq 1/1  to: AW284602  from: 1  to: 327

       2 ThrSerGluGluSerTrpThr 8
         |||| |||||||||||||||
      47 ACAICAGAAGAGTCATGGACT 57

Seq name: gb:est73:BE365425

Seq documentation block:
LOCUS       BE365425       406 bp    mRNA             EST       20 JUL 2000
DEFINITION  WS1_6_B02_g1 A002 Water stressed 1 (WS1) Sorghum bicolor cDNA,
            mRNA sequence.
ACCESSION   BE365425
VERSION     BE365425.1  GI:9305082
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 406)
  AUTHORS   Cordonnier Pratt,M. M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: water stressed plants
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
```

```
            Seq primer: PolyTMix
            High quality sequence start: 16
            High quality sequence stop: 380
            POLYA Yes.
FEATURES             Location/Qualifiers
     source          1..406
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Water-stressed I (WSI)"
                     /note="Organ: Mix of 5-week old plants on days 7 & 8 after
                     water was withheld; Vector: Lambda Zap; Site_1: XhoI;
                     Site_2: EcoRI; The library was made from poly A RNA in the
                     cloning vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT       97 a     93 c    151 g     65 t
ORIGIN


alignment_scores:
            Quality:     7.00             Length:       7
              Ratio:    1.000              Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_229_237 x BE463525

 Align seq 1/1  to: BE463525  from: 1  to: 406

       2 ThrSerGluGlnSerTrpThr 8
         ||||||||||||||||||||
     410 ACATCAGAAGAGTCATGGACT 430

Seq_name: gb_est50:AW679111

Seq_documentation_block:
LOCUS        AW679111       476 bp    mRNA             EST       19-JUL-2000
DEFINITION   WS1_22_G07.g1_A002 Water-Stressed I (WSI) Sorghum bicolor cDNA
             mRNA sequence.
ACCESSION    AW679111
VERSION      AW679111.1  GI:7552868
KEYWORDS     EST.
SOURCE       sorghum.
  ORGANISM   Sorghum bicolor
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
             clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE    1  (bases 1 to 476)
  AUTHORS    Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
             ,L.H.
  TITLE      An EST database from Sorghum: water-stressed plants
  JOURNAL    Unpublished (2000)
COMMENT      Contact: Cordonnier-Pratt MM
             Department of Botany
             The University of Georgia
             Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
             Tel: 706 542 1860
             Fax: 706 542 1805
             Email: mmpratt@uga.edu
             Sequences have been trimmed to exclude PolyA, vector and regions
             below Phred quality 16. The threshold for highest quality sequence
             is 20.
             Seq primer: T7
             High quality sequence start: 29
             High quality sequence stop: 459
             POLYA No.
FEATURES             Location/Qualifiers
     source          1..476
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Water-stressed I (WSI)"
                     /note="Organ: Mix of 5-week old plants on days 7 & 8 after
                     water was withheld; Vector: Lambda Zap; Site_1: XhoI;
                     Site_2: EcoRI; The library was made from poly A RNA in the
                     cloning vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT       122 a    105 c    146 g    111 t       2 others
ORIGIN


alignment scores:
            Quality:     7.00             Length:       7
              Ratio:    1.000              Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_229_237 x AW679111  ..

 Align seq 1/1  to: AW679111  from: 1  to: 476

       2 ThrSerGluGlnSerTrpThr 8
         ||||||||||||||||||||
     248 ACATCAGAAGAGTCATGGACT 268

Seq_name: gb_gss27:AZ474732

Seq_documentation_block:
LOCUS        AZ474732       554 bp    DNA              GSS       04-OCT-2000
DEFINITION   MMGGC1M0289H14 R, 10kb plasmid UUGC1M library Mus musculus genomic
             clone UUGC1M0289H14 R, DNA sequence.
ACCESSION    AZ474732
VERSION      AZ474732.1  GI:10631857
KEYWORDS     GSS.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciuromorphi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 554)
  AUTHORS    Dunn,D., Aoyagi,A., Barber,M., Beacorn,E., Duval,B., Hamil,C.,
             Islam,B., Longacre,S., Mahmood,M., Meenen,E., Pedersen,I., Reilly
             ,M., Ross,M., Ross,K., Stokes,K., Tingey,A., von Niederhausern,A.
             and Wright,D.,Weiss,R.
  TITLE      Mouse whole genome scaffolding with paired end reads from 10kb
             plasmid inserts
  JOURNAL    Unpublished (2000)
COMMENT      Contact:  Robert B. Weiss
             University of Utah Genome Center
             University of Utah
             Rm. 338, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
             84112, USA
             Tel: 801 585 5606
             Fax: 801 585 7177
             Email: ddunn@genetics.utah.edu
             Insert Length: 10000  Std Error: 0.00
             Plate: 0289  row: H  column: 14
             Seq primer: CACACAGGAAACAGCTATGACC
             Class: plasmid ends
             High quality sequence stop: 554.
FEATURES             Location/Qualifiers
     source          1..554
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0289H14"
                     /clone_lib "Mouse 10kb plasmid UUGC1M library"
                     /sex "Male"
                     /lab_host="E. Coli strain XL10 Gold, Tl resistant, F'"
                     /note "Vector: PW342nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/cmr/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size selected for a 9.5 to
```

10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWM42 (gi|4732114|gb|AF129072.1), a copy number inducible derivative of plasmid RI. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically competent E. coli XL10 Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 162 a 134 c 153 g 104 t 1 others
ORIGIN

Alignment scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

Alignment block:
US-09-126-945B-2_COPY_229_237 x AZ474732

Align seq 1/1 to: AZ474732 from: 1 to: 554

1 SerThrSerGluGluSerTrp 7
241 TCGACCTCCGAGGAAAGCTGG 261

Seq name: gb:est76:BE597371

Seq documentation block:
LOCUS BE597371 579 bp mRNA EST 18-AUG-2000
DEFINITION PTI 72_F10_q1_A002 Pathogen-Induced 1 (PTI) Sorghum bicolor cDNA, mRNA sequence.
ACCESSION BE597371
VERSION BE597371.1 GI:9852444
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 579)
AUTHORS Cordonnier-Pratt,M. M., Gingle,A., Dean,R., Sudman,M. and Pratt ,L.H.
TITLE An EST database from Sorghum: pathogen induced plants
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: PolyTMix
High quality sequence start: 44
High quality sequence stop: 574
POLYA No.
FEATURES Location/qualifiers:
source 1..579
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Pathogen-induced 1 (PTI)"
/note="Organ: Anthracnose-infected leaves from two-week-old sorghum plants 48 hr after inoculation; Vector: pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2: EcoRI; Two week old sorghum plants (BTX 623 cultivar) were infected with pathogen (isolate: PEM1.1 = Colletotrichum graminicola, which is a sorghum isolate). RNA was prepared from infected leaves harvested from 45 seedlings 48 hours after inoculation. Note: young seedlings (2 weeks old) exhibit juvenile resistant reaction, which is an incompatible interaction. As they grow older (4 weeks or older), plants regain susceptibility to anthracnose disease. The library was made from poly A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision. WARNING: While most or all ESTs are expected to derive from the host plant, no effort was made to eliminate ESTs deriving from the pathogen."
BASE COUNT 133 a 139 c 177 g 130 t
ORIGIN

Alignment scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

Alignment block:
US-09-126-945B-2_COPY_229_237 x BE597371

Align seq 1/1 to: BE597371 from: 1 to: 579

2 ThrSerGluGluSerTrpThr 8
429 ACAICAGAAGAGTCATGGACT 449

Seq name: gb:est89:BE574485

Seq documentation block:
LOCUS BE574485 713 bp mRNA EST 12-DEC-2000
DEFINITION 602132422F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4271765 5', mRNA sequence.
ACCESSION BE574485
VERSION BE574485.1 GI:11648197
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 713)
AUTHORS [illegible]
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLCM1084 row: l column: 06
High quality sequence stop: 620.
FEATURES Location/qualifiers
source 1..713
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4271765"
/clone_lib="NIH_MGC_81"
/tissue_type="THP1(PI phage resistant)"
/note="Organ: muscle (skeletal); Vector: pINCY [Incyte]; Site_1: SfiI (ggccgcctcgggcc); Site_2: SfiI (ggccattatgggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5' CACGGCCATTATGGCC 3' and 3' adaptor sequence: 5' ATTCTACAGGCCGAGGCGGCCGACATG-d(30)BN 3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.5 Kb (range 1.0-3.0 Kb). 1.75 colonies contained inserts by PCR. This library was enriched for full length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

```
BASE COUNT      208 a    132 c    197 g    176 t
ORIGIN


alignment_scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_229_237 x BF574485  ..

 Align seq 1/1  to: BF574485  from: 1  to: 713

       1 SerThrSerGluGluSerTrp 7
         |||||||||||||||||||||
     639 AGTACCTCGGAAGAATCATGG 659

seq_name: gb_est89:BF582587

seq_documentation_block:
LOCUS        BF582587      744 bp    mRNA            EST       11-DEC-2000
DEFINITION  602074643F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:4211698 5',
            mRNA sequence.
ACCESSION   BF582587
VERSION     BF582587.1  GI:11619950
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 744)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9781  row: e  column: 11
            High quality sequence stop: 669.
FEATURES             Location/Qualifiers
     source          1..744
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4211698"
                     /clone_lib "NCI_CGAP_Li9"
                     /lab_host "DH10B (T1 phage-resistant)"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: oligo dT.
                     Average insert size 1.9 kb. Constructed by Life
                     Technologies. Note: this is a NCI CGAP Library."
BASE COUNT      182 a    165 c    201 g    196 t
ORIGIN


alignment_scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_229_237 x BF542587  ..

 Align seq 1/1  to: BF542587  from: 1  to: 744

       4 SerGluGluSerTrpThrAsp 9
         |||||||||||||||||||||
     592 AGTGAAGAATCTTGGACTGAC 612

seq_name: gb_est89:BE040978

seq_documentation_block:
LOCUS        BE040978      1004 bp    mRNA            EST       07-JUN-2000
DEFINITION  OF15F03.3F Oryza sativa cDNA 5', mRNA sequence.
ACCESSION   BE040978
VERSION     BE040978.1  GI:8336601
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 1004)
  AUTHORS   Bohnert,H.J., Borchert,C., Brazille,S., Brooks,J., Eaton,M., Ferrea
            ,B., Kawasaki,S., McCollough,A., Michalowski,C.B., Palacio,C.,
            Scara,G., Wheeler,M. and Zepeda,G.R.
  TITLE     Functional Genomics of Plant Stress Tolerance
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Michalowski,C.B.
            University of Arizona
            Bio Sciences West room 518, Tucson, AZ 85721, USA
            Tel: 520-621-7982
            Fax: 520-621-1697
            Email: cbm@ag.arizona.edu
            Best blastx match: 'gb|AAD34675.1|AC006341 3 (AC006341) ESTs
            gb|BE15498: gb|BE37515: g. . . 155 7e-47'. An open reading frame
            exists.
FEATURES             Location/Qualifiers
     source          1..1004
                     /organism "Oryza sativa"
                     /strain "Pokkali"
                     /db_xref "taxon:4530"
                     /clone_lib "OF"
                     /tissue_type "entire plant"
                     /dev_stage="2 weeks"
                     /note="1 week 150mM NaCl"
BASE COUNT      251 a    256 c    221 g    277 t
ORIGIN


alignment_scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_229_237 x BE040978/rev  ..

 Align seq 1/1  to reverse of: BE040978  from: 1  to: 1004

       2 ThrSerGluGluSerTrpThr 8
         |||||||||||||||||||||
     466 ACTTCCGAGGAATCTTGGACA 446

seq_name: gb_est98:BG243672

seq_documentation_block:
LOCUS        BG243672     1019 bp    mRNA            EST       13-FEB-2001
DEFINITION  602527720F1 NCI_CGAP_Mus1 Mus musculus cDNA clone IMAGE:4486407 5',
            mRNA sequence.
ACCESSION   BG243672
VERSION     BG243672.1  GI:12756487
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1019)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
```

```
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10329 row: k column: 16
            High quality sequence start: 2
            High quality sequence stop: 637.
FEATURES             Location/Qualifiers
     source          1..1019
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4486407"
                     /clone_lib="NCI_CGAP_Mam1"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="4 months, virgin"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally. Primer: oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT       261 a    246 c    265 g    257 t
ORIGIN


Alignment scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-126-945B-2_COPY_229_237 x BG243672/seq   ..

 Align seq 1/1  to reverse of: BG243672  from: 1  to: 1019

        2 ThrSerGluGluSerTrpThr 8
          ||||||||||||||||||||
      697 ACCAGTGAAGAATCCTGGACC 677

Seq name: gb:gss16:AQ752331

Seq documentation block:
LOCUS       AQ752331      1221 bp    DNA              GSS        19-JUL-1999
DEFINITION  HS_5565_B2_D09_F7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate 1141 Col 18 Row H, DNA sequence.
ACCESSION   AQ752331
VERSION     AQ752331.1  GI:5549489
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1221)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-4887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@roswell.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.chori.org/ordering.htm)
            or from Research Genetics (info@resgen.com). BAC end Web Server:
            http://www.hgsc.washington.edu
            Plate: 1141 row: H column: 18
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 1221.
FEATURES             Location/Qualifiers
     source          1..1221
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate 1141 Col 18 Row H"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="Male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoKI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT       278 a    242 c    350 g    239 t      12 others
ORIGIN


Alignment scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-126-945B-2_COPY_229_237 x AQ752331   ..

 Align seq 1/1  to: AQ752331  from: 1  to: 1221

       1 SerThrSerGluGluSerTrp 7
         |||||||||||||||||||
     858 AGCACCAGTGAAGAATCCTGG 878

Seq name: gb:est2:AA086599

Seq documentation block:
LOCUS       AA086599      148 bp    mRNA             EST        15-FEB-1997
DEFINITION  mn91t05.r1 Stratagene mouse Tcell 937411 Mus musculus cDNA clone
            IMAGE:559457 5', mRNA sequence.
ACCESSION   AA086599
VERSION     AA086599.1  GI:1629807
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 148)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:332249
            Seq primer:  28ml3 rev1 ET from Amersham
```

```
            High quality sequence stop: 133.
FEATURES             Location/Qualifiers
     source          1. .148
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5514457"
                     /clone_lib="Stratagene mouse Tcell 937811"
                     /tissue_type="Tcell"
                     /dev_stage="M30 CD4+ cells"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="organ: blood; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. M30 CD4+ cells. Average insert size: 1.0 kb;
                     Uni-ZAP XR Vector; 5' adaptor sequence: 5' GAATTCGGCACGAG
                     3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT       28 a      41 c      38 g      41 t
ORIGIN


alignment_scores:
            Quality:    6.00             Length:      6
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_229_237 x AA086599/rev

 Align seq 1/1  to reverse of. AA086599  from: 1  to: 148

        3 SerGluGluSerTrpThr 8
          ||||||||||||||||||
       79 TCACAACAGAGTTGGACT 53

seq_name: qb_est92:BF772068

seq_documentation_block:
LOCUS        BF772068      173 bp    mRNA            EST       12-JAN-2001
DEFINITION  CM4-TT0041-051200-477-b01 TT0041 Homo sapiens cDNA, mRNA sequence.
ACCESSION    BF772068
VERSION      BF772068.1  GI:12119968
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 159)
  AUTHORS    Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
             Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
             Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
             Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
             ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
             Simpson,A.J.
  TITLE      Shotgun sequencing of the human transcriptome with ORF expressed
             sequence tags
  JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE    20202663
COMMENT      Contact: Simpson A.J.G.
             Laboratory of Cancer Genetics
             Ludwig Institute for Cancer Research
             Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
             Brazil
             Tel: +55-11-2704922
             Fax: +55-11-2707001
             Email: asimpson@ludwig.org.br
             This sequence was derived from the FAPESP/LICR Human Cancer Genome
             Project. This entry can be seen in the following URL
             (http://www.ludwig.org.br/cgi-perl/getbyml2.pl?url=CM4&id=CM4-TT0041-
             051200-477-b01&t=2000-12-05&t4=1)
             Seq primer: puc 18 forward
             High quality sequence start: 19
             High quality sequence stop: 126.
FEATURES             Location/Qualifiers
     source          1. .159
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="TT0041"
                     /dev_stage="Adult"
                     /note="Organ: epid_tumor; Vector: pucl8; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       44 a      36 c      44 g      35 t
ORIGIN


alignment_scores:
            Quality:    6.00             Length:      6
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_229_237 x BF772068   ..

 Align seq 1/1  to: BF772068  from: 1  to: 159

       1 PerTlePerGluGluSer 6
         ||||||||||||||||||
      59 TCTACATCAGAAGAATCT 76

seq_name: qb_qss5:FR0004420

seq_documentation_block:
LOCUS        FR0004420      159 bp    DNA             GSS       27-FEB-1997
DEFINITION  F.rubripes GSS sequence, clone 047P21aF9, genomic survey sequence.
ACCESSION    Z88203
VERSION      Z88203.1  GI:1885115
KEYWORDS     GSS; genome survey sequence.
SOURCE       Takifugu rubripes.
  ORGANISM   Takifugu rubripes
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
             Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
             Tetraodontidae; Takifugu.
REFERENCE    1  (bases 1 to 159)
  AUTHORS    Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrania,Y.,
             Williams,G. and Brenner,S.
  TITLE      Direct Submission
  JOURNAL    Submitted (18-FEB-1997) MRC Human Genome Mapping Project Resource
             Centre Hinxton, Cambridge, CB10 1SB, Email: bioinfo@hgmp.mrc.ac.uk
COMMENT      Vector: m13mp18
             V_type: phage
             PRIMER: M13
             DESCR:
             One pass dye terminator sequencing of cosmid cloned genomic
             sequence.
FEATURES             Location/Qualifiers
     source          1. .159
                     /organism "Takifugu rubripes"
                     /db_xref="taxon:31033"
                     /clone_lib "cosmid 047P21"
                     /clone "047P21aF9"
BASE COUNT       54 a      40 c      31 g      34 t
ORIGIN


alignment_scores:
            Quality:    6.00             Length:      6
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_229_237 x FR0004420/rev   ..
```

Align seq 1/1 to reverse of: FR0004420 from: 1 to: 159

```
      2 ThrSerGluGluSerTrp 7
        ||||||||||||||||||
     52 ACAICTGAAGAATCCTGG 35
```

```
CM of: US-09-126-945B-2_COPY_234_242 to: GenEmbl:*   -out_format : pir

Date: Sep 14, 2001  9:49 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL trame=_p2n.model -DEV xlp
 -Q /cgn2_6/PTO/tmp/US09_126_945B_2_COPY_234_242.rge_1_7594.qry -DB GenEmbl
 -QFMT fastap -SUFFIX olip2n.rge -GAPOP 4.500
 -GAPEXT 0.050 -MINMATCH=0.100 -LOOPOL 0.000 -LOOPEXT 0.000
 -QGAPOP 4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT 60.000
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT 60.000
 -DELOP 6.000 -DELEXT=7.000 -START 1 -MATRIX=olip
 -TRANS human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE quality
 -THR_MIN 1 -ALIGN 15 -MODE=LOCAL -OUTFMT pir -NORM ext -MINLEN 9
 -MAXLEN 2000000000 -USER=US09126945_2_COPY_1_7594 -NCPU 6
 -ICPU 3 -LONGDOC -NO_XLPXY -WAIT -THREADS 1

Search information block:
Query: US-09-126-945B-2_COPY_234_242
Query length: 9
Database: GenEmbl:*
Database sequences: 1844157
Database length: -856060004
Search time (sec): 6503.700000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000


score list:
Sequence                Strd Orig    %Score   EScore Len  ! Documentation ..
qb_pat1:AX048148          +     9.00  192.32  0.0178    1894  ! AX048148 Sequence 1 from Paten
qb_pr4:AF071538           +     9.00  192.32  0.0178    1894  ! AF071538 Homo Sapiens Ets tran
qb_pr1:AB031549           +     9.00  192.41  0.0178    1898  ! AB031549 Homo sapiens PSE mRNA
qb_pat1:AX048159          +     9.00  188.31  0.0298    4417  ! AX048159 Sequence 12 from Pate
qb_pr6:AL157272           -     9.00  160.97  0.0622  150660  ! AL157272 Homo DNA seq.e ...
qb_in2:CEC46C2            +     7.00  125.66   92.05     41115  ! Z68296 Caenorhabditis elegans
qb_pr2:AC008530           +     7.00  120.12  187.34     80095  ! AC008530 Homo sapiens chromos
qb_pr4:AC022428           -     7.00  119.67  198.51     94890  ! AC022428 Homo sapiens chromos
qb_htq1:AC008879          -     7.00  117.71  255.12    124682  ! AC008879 Homo sapiens chromo
qb_pr3:AC016591           +     7.00  117.02  278.86    147362  ! AC016591 Homo sapiens chromo
qb_htq5:AC016212          -     7.00  115.92  321.22    168217  ! AC016212 Homo sapiens clone
qb_pr8:HS125B2            -     7.00  115.34  345.64    178513  ! Z98949 Human DNA sequence fr
qb_htq20:AL158154         -     7.00  115.12  355.86    179104  ! AL158154 Homo sapiens chromo
qb_htq16:AC079219         +     7.00  115.04  359.33    181088  ! AC079219 Mus musculus chromo
qb_htq5:AC016583          -     7.00  115.00  361.29    182082  ! AC016583 Homo sapiens chromo
qb_pr5:AL133417           -     7.00  114.91  365.17    184212  ! AL133417 Human DNA sequence
qb_htq20:AL355380         -     7.00  114.83  368.98    186302  ! AL355380 Homo sapiens chromo
qb_htq14:AC067751         +     7.00  113.56  434.43    222537  ! AC067751 Homo sapiens chromo
qb_htq12:AC044804         +     7.00  112.64  489.08    253123  ! AC044804 Mus musculus chromo
qb_htq20:AL353740         -     7.00  112.43  502.46    260714  ! AL353740 Homo sapiens chromo
qb_htq25:HSBA22B4         +     7.00  111.45  569.12    298575  ! AL080244 Homo sapiens chromo
qb_sts2:G24671            +     6.00  139.68   15.24       258  ! G24671 human STS WI 15721, sequ
qb_sts2:G23557            +     6.00  137.23   26.86       363  ! G23557 human STS WI-30916, sequ
qb_ov:POUMTCBF            +     6.00  132.21   39.73       732  ! D50025 Saw shark Mitochondrial
qb_pl3:CNS01DED           -     6.00  131.75   42.12       780  ! AL116717 Botrytis cinerea strai
qb_sts1:CNS01LB9          +     6.00  131.46   43.75       813  ! AL149630 Anopheles gambiae STS
qb_sts1:CNS071A8          +     6.00  130.58   48.47       919  ! AL424662 13 end of clone XA20AA
qb_sts1:CNS06T5F          -     6.00  130.43   49.45       939  ! AL413121 13 end of clone AW0AA0
qb_sts1:CNS071V0          +     6.00  129.78   54.29      1028  ! AL125443 clone BA0AB005H02 of
qb_ov:AF006222            +     6.00  129.66   55.11      1045  ! AF006222 Chemoscopis rubrirost
qb_ov:AF006243            +     6.00  129.66   55.11      1045  ! AF006243 Nephelornis oneilli c
qb_pat2:I08775            -     6.00  128.46   64.30      1236  ! I08775 Sequence 8 from Patent
qb_pat2:I08920            -     6.00  128.46   64.30      1236  ! I08920 Sequence 43 from Patent
qb_sts2:G26845            -     6.00  128.22   66.30      1278  ! G26845 human STS UTK 9857, seq
qb_pr7:BC004261           +     6.00  127.82   69.77      1351  ! BC004261 Homo sapiens, clone M
qb_rol:AF055065           +     6.00  127.80   69.91      1354  ! AF055065 Rattus norvegicus sig
qb_pl3:CTAJ5082           +     6.00  127.77   70.20      1360  ! AJ005082 Cyanopsis tetragonolo
qb_ov:XELFIBRIL           +     6.00  126.77   79.02      1564  ! M28874 X.laevis fibrillarin (X
qb_sts2:G66917            -     6.00  125.87   89.61      1774  ! G66917 cnpgaad67 pr17-1 Human
qb_rol:MMEP6682           +     6.00  125.61   92.58      1838  ! UP6682 Mus musculus immunos
qb_rol:AF148879           +     6.00  124.91  101.38      2029  ! AF148879 Mus musculus strai
qb_pat2:E28447            +     6.00  124.86  102.03      2043  ! E28447 Novel peptide inhibi
qb_rol:AF041035           +     6.00  124.86  102.03      2043  ! AF041035 Mus musculus K-Ric
qb_rol:AF148880           +     6.00  124.70  104.04      2087  ! AF148880 Mus musculus strai
qb_rol:AF148878           +     6.00  124.69  104.27      2092  ! AF148878 Mus musculus strai

Seq_name: qb_pat1:AX048148

Seq_documentation_block:
LOCUS       AX048148     1894 bp    DNA             PAT       15-DEC-2000
DEFINITION  Sequence 1 from Patent WO0070092.
ACCESSION   AX048148
VERSION     AX048148.1  GI:11876967
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1894)
  AUTHORS   Kaufmann,J. and Xin,H.
  TITLE     Expression of ets domain proteins in cancer
  JOURNAL   Patent: WO 0070092-A 1 23-NOV-2000;
            Chiron Corporation (US)
FEATURES             Location/Qualifiers
     source          1..1894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             416..1423
                     /note="human JKETS."
                     /codon_start=1
                     /protein_id="AAL19002.1"
                     /db_xref="GI:11876968"
                     /translation="MGSASPGLSSVSPSHLLLPFDLVSRIGLEKAAAGAVGLEPRLWS
                     PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAPGASSREPPPFEPEQCPVTDSQAPA
                     GSLDLVPGGLELLEHSLEQVQSMVVGEVLKDIEIACKLLNLTADPMDWSPSNVQFWLL
                     WTEHGYKEIPMGKAFQELLACKEDCAMSLLQLEKGKSPIGGDVLHAHLETWKSAAWMKER
                     TSPGAIHTVASISFESWTDSEVDSSCSGQICDLWQETKELLKLDSYCKEIRWLNFEE
                     NFPIIGSA_VAELWCDFEKEIAMNYQELSPSIE_YYEECIIRKPTIPQMLVY_EVHP
                     I"
BASE COUNT      368 a    653 c    571 g    302 t
ORIGIN


alignment_scores:
            Quality:    9.00              Length:      9
              Ratio:   1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_234_242 x AX048148  ..

 Align seq 1/1  to: AX048148  from: 1  to: 1894

       1 SerTrpThrAspSerGluValAspSer 9
         ||||||||||||||||||||||||||
    1115 AGCTGGACCGACAGCGAGGTGGACTCA 1141

Seq_name: qb_pr4:AF071538

Seq_documentation_block:
LOCUS       AF071538     1894 bp    mRNA            PRI       10-JAN-2000
DEFINITION  Homo Sapiens Ets transcription factor ESE (ESE) mRNA, complete
            cds.
ACCESSION   AF071538
VERSION     AF071538.1  GI:4007417
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo Sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1894)
  AUTHORS   Oettgen,P., Finger,E., Sun,Z., Akbarali,Y., Thamrongsak,U.,
```

```
            Boltax,I., Crall,F., Dube,A., Weiss,A., Brown,L., Quinn,G., Kas,K.,
            Endress,G., Kunsch,C. and Libermann,T.A.
  TITLE     PDEF, a novel prostate epithelial specific ets transcription
            factor, interacts with the androgen receptor and activates
            prostate specific antigen gene expression
  JOURNAL   J. Biol. Chem. 275 (2), 1216-1225 (2000)
  MEDLINE   20092894
REFERENCE   2  (bases 1 to 1894)
  AUTHORS   Libermann,T.A., Oettgen,P., Kunsch,C. and Finger,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (10 JUN 1998) Medicine, Beth Israel Deaconess Medical
            Center, 330 Brookline Ave., Boston, MA 02215, USA
FEATURES             Location/Qualifiers
     source          1..1894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6p21.3"
                     /tissue_type="prostate"
     gene            1..1894
                     /gene="PDEF"
     CDS             416..1423
                     /gene="PDEF"
                     /note="prostate epithelium specific"
                     /codon_start=1
                     /product="Ets transcription factor PDEF"
                     /protein_id="AAC95296.1"
                     /db_xref="GI:4667416"
                     /translation="MGSAGFGLSAVGPSHLLLLPPQIVSKEGLEKAAAGAVGLEKRDWS
                     PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAPGASSREEPPEEPEQCPVIDSQAPA
                     GSLDLVPGGGTLGDHSLEQVQSMVVGPVLKDLDIYACKLLNLTADEMDWSPSNVQKWLL
                     WTEHQYKLPPMGKAFQELAGKELCAMSEEQFRQPSRLGGAVLPAHLDTWRSAAWMEEP
                     TSPGATHYCASTSEEESWTDSEVDSSCSGQFCHLWQELEELLLKPHEYGPLFWELNKEE
                     GIFKIEDSAQVARLWGIRKNKPAMNYDKLSRSIRQYYKKGIIRKPDTSQRLVYQFVHP
                     I"
BASE COUNT       368 a    653 c    571 g    302 t
ORIGIN


alignment scores:
            Quality:    9.00              Length:      9
              Ratio:   1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2 COPY_234_242 x AF071538   ..

 Align seq 1/1  to: AF071538  from: 1  to: 1894

       1 SerTrpThrAspSerGluValAspSer 9
         |||||||||||||||||||||||||||
    1115 AGCTGGACCGACAGCGAGGTGGACTCA 1141

seq name: qb pri:AB031549

seq documentation block:
LOCUS       AB031549     1898 bp    mRNA            PRI       20 JAN 2000
DEFINITION  Homo sapiens PSE mRNA for prostate ets, complete cds.
ACCESSION   AB031549
VERSION     AB031549.1  GI:6721497
KEYWORDS    prostate ets.
SOURCE      Homo sapiens adult male prostate epithelium prostate cancer
            cell line:PC 3 cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Yamada,N., Tamai,Y., Miyamoto,H. and Nozaki,M.
  TITLE     Cloning and expression of the mouse Pse gene encoding a novel Ets
            family member
  JOURNAL   Gene 241 (2), 267-274 (2000)
  MEDLINE   20137504
REFERENCE   2  (bases 1 to 1898)
  AUTHORS   Nozaki,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25 AUG 1999) to the DDBJ/EMBL/GenBank databases. Masami
            Nozaki, Osaka University, Institute for Microbial Diseases; 3-1,
            Yamadaoka, Suita, Osaka 565-0871, Japan
            (E-mail:nozaki@biken.osaka-u.ac.jp, Tel:81-6-879-8338,
            Fax:81-6-879-8339)
FEATURES             Location/Qualifiers
     source          1..1898
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="PC-3"
                     /cell_type="prostate cancer"
                     /chromosome="6"
                     /dev_stage="adult"
                     /map="6p21.2-21.3"
                     /sex="male"
                     /tissue_type="prostate epithelium"
                     /vector="pUC18"
     gene            406..1413
                     /gene="PSE"
     CDS             406..1413
                     /gene="PSE"
                     /codon_start=1
                     /product="prostate ets"
                     /protein_id="BAA89543.1"
                     /db_xref="GI:6721498"
                     /translation="MGSASPGLSAVSPSHLLLLPPQIVSKEGLEKAAAGAVGLEKRDWS
                     PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAPGASSREEPPEEPEQCPVIDSQAPA
                     GSLDLVPGGGTLGDHSLEQVQSMVVGPVLKDLDIYACKLLNLTADEMDWSPSNVQKWLL
                     WTEHQYKLPPMGKAFQELAGKELCAMSEEQFRQPSRLGGAVLPAHLDTWRSAAWMEEP
                     TSPGATHYCASTSEEESWTDSEVDSSCSGQFCHLWQELEELLLKPHEYGPLFWELNKEE
                     GIFKIEDSAQVARLWGIRKNKPAMNYDKLSRSIRQYYKKGIIRKPDTSQRLVYQFVHP
                     I"
BASE COUNT       380 a    651 c    567 g    400 t
ORIGIN


alignment scores:
            Quality:    9.00              Length:      9
              Ratio:   1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2 COPY_234_242 x AB031549   ..

 Align seq 1/1  to: AB031549  from: 1  to: 1898

       1 SerTrpThrAspSerGluValAspSer 9
         |||||||||||||||||||||||||||
    1105 AGCTGGACCGACAGCGAGGTGGACTCA 1131

seq name: qb pat1:AX048159

seq documentation block:
LOCUS       AX048159     3317 bp    DNA             PAT       15-DEC-2000
DEFINITION  Sequence 12 from Patent WO0070092.
ACCESSION   AX048159
VERSION     AX048159.1  GI:11876975
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3317)
  AUTHORS   Kaufmann,I. and Xin,H.
  TITLE     Expression of ets-domain proteins in cancer
  JOURNAL   Patent: WO 0070092 A 12 23-NOV-2000;
            Chiron Corporation (US)
FEATURES             Location/Qualifiers
     source          1..3317
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

```
                     /note "GSEF encoding sequence with promoter (Figs. 2A 2B)"
     promoter        1. .1480
                     /note="Promoter"
     CDS             1815. .2822
                     /note="GSEF coding sequence"
                     /codon_start 1
                     /protein_id "CAC19004.1"
                     /db_xref "GI:11876976"
                     /translation="MGSASPGISSVSESHEIIEPDIVSKFLEKAAACAVCLEPPPWS
                     PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAPCASSREEPPPFEPEQCPVIDSQAPA
                     GSLDLVPGGLETLEEHSLEQVQSMVVTEVLKTLETACKLLNLTAETPAWSPSNVQKWLL
                     WIEHQYRLPDMGKAFLFLAGEFLAMSFQLFKQWSFPLTVVFHAHLFLWESAAWMKEP
                     TSPGATHYCASTSFFSWFLFSFVDSSCSQQFTLWLFLKFLFLKPHSYCPFFRWLNAAA
                     GIFFTFDSAQVAAFWGTFRNPFAMNYFFSGFIFQAVSFLQKFVVLIVSLIVS
                     1"
BASE COUNT       710 a   1026 c    970 g    611 t
ORIGIN


alignment_scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_234_242 x AX048159     .

 Align seq 1/1  to: AX048159  from: 1  to: 3317

        1 SerTrpThrAspSerGluValAspSer 9
          ||||||||||||||||||||||||||
     2514 AGCTGGACCGACAGCGAGGTGGACTCA 2540

seq_name: qb_pr6:AL157372

seq_documentation_block:
LOCUS        AL157372   150660 bp    DNA             PRI       29-NOV-2000
DEFINITION   Human DNA sequence from clone RP11-375E1 on chromosome 6, complete
             sequence.
ACCESSION    AL157372
VERSION      AL157372.18  GI:11493236
KEYWORDS     HTG.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 150660)
  AUTHORS    Skuce,C.
  TITLE      Direct Submission
  JOURNAL    Submitted (29-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
             CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
             requests: clonerequest@sanger.ac.uk
COMMENT      On Nov 30 2000 this sequence version replaced gi:11512772.
             During sequence assembly data is compared from overlapping clones.
             Where differences are found these are annotated as variations
             together with a note of the overlapping clone name. Note that the
             variation annotation may not be found in the sequence submission
             corresponding to the overlapping clone, as we submit sequences with
             only a small overlap as described above.
             This sequence has been finished according to sequence map criteria
             as follows.  An attempt is made to resolve all sequencing problems,
             such as compressions and repeats, but not necessarily within known
             annotated repeat sequence elements.  Where the sequence is
             ambiguous, there is an annotation using the 'unsure' feature key
             The following abbreviations are used to associate primary accession
             numbers given in the feature table with their source databases.
             Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
             on the WORMPEP database can be found at
             http://www.sanger.ac.uk/Projects/Wormpep/ This sequence
             was generated from part of bacterial clone contigs of human
             chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
             Group.  Further information can be found at
             http://www.sanger.ac.uk/HGP/Chr6/
             RP11-375E1 is from the library RPCI-11.2 constructed at the Roswell
             Park Cancer Institute by the group of Pieter de Jong. For further
             details see http://bacpac.med.buffalo.edu/
             VECTOR: pBACe3.6
             IMPORTANT: This sequence is not the entire insert of clone
             RP11-375E1 it may be shorter because we sequence overlapping
             sections only once, except for a 100 base overlap.
             The true left end of clone RP11-375E1 is at 1 in this sequence. The
             true left end of clone RP3-391O22 is at 150561 in this sequence.
             The true right end of clone RP11-302L10 is at 113044 in this
             sequence.
FEATURES             Location/Qualifiers
     source          1. .150660
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /chromosome "6"
                     /clone "RP11-375E1"
                     /clone_lib "RPCI-11.2"
BASE COUNT    45470 a   38887 c   38740 g   37563 t
ORIGIN


alignment_scores:
            Quality:    9.00              Length:       9
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_234_242 x AL157372/rev     ..

 Align seq 1/1  to reverse of: AL157372  from: 1  to: 150660

        1 SerTrpThrAspSerGluValAspSer 9
          ||||||||||||||||||||||||||
   124664 AGCTGGACCGACAGCGAGGTGGACTCA 124638

seq_name: qb_in2:CEC46C2

seq_documentation_block:
LOCUS        CEC46C2     41115 bp    DNA             INV       25-OCT-2000
DEFINITION   Caenorhabditis elegans cosmid C46C2, complete sequence.
ACCESSION    Z68296
VERSION      Z68296.1  GI:1130612
KEYWORDS     HTC; Serine/threonine protein kinase; Transport protein.
SOURCE       Caenorhabditis elegans.
  ORGANISM   Caenorhabditis elegans
             Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
             Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE    1  (bases 1 to 41115)
  AUTHORS    none.
  TITLE      Genome sequence of the nematode C. elegans: a platform for
             investigating biology. The C. elegans Sequencing Consortium
  JOURNAL    Science 282 (5396), 2012-2018 (1998)
  MEDLINE    99069618
  REMARK     The C.elegans Sequencing Consortium.
             Erratum:[[published errata appear in Science 1999 Jan
             1;283(5398):35 and 1999 Mar 26;283(5410):2103 and 1999 Sep
             3;285(5433):1493]]
REFERENCE    2  (bases 1 to 41115)
  AUTHORS    McMurray,A.A.
  TITLE      Direct Submission
  JOURNAL    Submitted (20-DEC-1995) Nematode Sequencing Project, Sanger Centre,
             Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
             Washington University, St. Louis, MO 63110, USA. E-mail:
             jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT      Coding sequences below are predicted from computer analysis, using
             predictions from GeneFinder (P. Green, U. Washington), and other
             available information.
             Current sequence finishing criteria for the C. elegans genome
             sequencing consortium are that all bases are either sequenced
             unambiguously on both strands, or on a single strand with both a
             dye primer and dye terminator reaction, from distinct subclones.
             Exceptions are indicated by an explicit note.
```

```
            IMPORTANT: This sequence is not the entire insert of clone C46C2.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone C46C2 is at 1 in this sequence. The true
            right end of clone C46C2 is at 104 in
            sequence AL032642.
            The start of this sequence (1. .104) overlaps with the end of
            sequence Z68295.
            The end of this sequence (41012. .41115) overlaps with the start of
            sequence AL032642.
            For a graphical representation of this sequence and its analysis
            see: http://www.ebi.ac.uk/cgi-bin/emblfetch?Z68296+1+41115+name
            name C46C2.
FEATURES                Location/Qualifiers
     source             1. .41115
                        /organism "Caenorhabditis elegans"
                        /db_xref "taxon:6239"
                        /chromosome "IV"
                        /clone "C46C2"
     gene               join(1. .54,256. .446,589. .707,870. .1114,2461. .2648,
                        2993. .3485,3538. .3695,3840. .3941,4023. .4319)
                        /gene "C46C2.4"
     CDS                join(1. .54,256. .446,589. .707,870. .1114,2461. .2648,
                        2993. .3485,3538. .3695,3840. .3941,4023. .4319)
                        /gene "C46C2.4"
                        /note "contains similarity to Pfam domains PF00561
                        (alpha/beta hydrolase fold) Score 40.8, E value 1e-05,
                        N 1"
                        /codon_start 1
                        /protein_id "CAA92590.1"
                        /db_xref "GI:3874924"
                        /db_xref "SPTREMBL:Q18656"
                        /translation "MCSICSNSYLNEIQRYLYSHEGSLNNEKIALLYVSQQNTLNLN
                        ILKQGPFVQAAEEFGASMFALEHREYGNSKPRSMNLISKDLKYLKSSEAVQDTISFIN
                        YSNKKFNMNPCVKRSKYNEQVENSILKFCQQECFQKIETCFNYLRAAMNTIQCRKELS
                        DQFKLEPRLNTQNPNSNEIQLFYLYVIAPFQQTVQFHRQQRFFKNSKYAKEQGPNFLC
                        IGQEGREDPNSIREDVFAVVFKAQKFGATVYVFEHREYGDSNVGDNSDLSKISSLQME
                        YDLAETIKEENLKTNTSNPWITFGGSYSGMLSAWMRETFHEFVVGAVASSAPILAKTD
                        FYEYIMVVEDVFRRYDIGCYNATKKGFLEIQKMFETIEQRQKLSKIFPSYFALKNNFS
                        ETRKHEFLLDLADPEETSVQYAGYGSGAFAKTSKKLKNKFKGVFDGQNEDLISLENIG
                        LWMWQTCTEFGFYQQSTDTGNSIFGNVIVSYFVQQCMDLFGNNYCRATIDKQVSRFNHK
                        YDGTYEFNATNVVFLNGDADPWSPLGLKNSTDPSVVSFLINGTSHCVDMYSFTETDDLP
                        DLEIAREIVDNIFEWLNCTLTFEPCCANYFEALNCIAVLIVYYVII"
     gene               complement(join(4878. .5195,5289. .5492,5628. .5819))
                        /gene "C46C2.5"
     CDS                complement(join(4878. .5195,5289. .5492,5628. .5819))
                        /gene "C46C2.5"
                        /note "cDNA EST yk579a1.3 comes from this gene;
                        cDNA EST yk579a1.5 comes from this gene"
                        /codon_start 1
                        /protein_id "CAB54197.1"
                        /db_xref "GI:5824407"
                        /db_xref "SPTREMBL:Q9U3M0"
                        /translation "MASICFTIPNQEMEFEETEVIPNCTVSVDAFLPSLESCLENC
                        LSKNCLTLTYSFLTLDRETETLQNSLNNVIVAIELLSELLELLKLSGSLRKSCD
                        SSMIFAEYLLDGIVFTGDNIYGEAFNCLSKEFTLVQTKYNSFIQENLESLFMSCAEEI
                        TLNNQSIDILEHLKPVSPCVAGFLASLELPNSFTAVFAVCVFLFTFLNSRFTLSGI
                        KENKMKLEFNKEKYYSYLN"
     gene               join(14510. .14554,15442. .15520,15825. .15894,16489. .16573,
                        16911. .17347,17403. .17704,17774. .18099,18147. .19391,
                        19680. .20378,20429. .21194,21291. .21732,22299. .22772,
                        23560. .23612,23668. .23872,24447. .24776)
                        /gene "C46C2.1"
     CDS                join(14510. .14554,15442. .15520,15825. .15894,16489. .16573,
                        16911. .17347,17403. .17704,17774. .18099,18147. .19391,
                        19680. .20378,20429. .21194,21291. .21732,22299. .22772,
                        23560. .23612,23668. .23872,24447. .24776)
                        /gene "C46C2.1"
                        /note "similarity to Mouse A-RAF proto-oncogene
                        serine/threonine protein kinase (SW:KRAA_MOUSE), contains
                        similarity to Pfam domains PF00069 (Eukaryotic protein
                        kinase domain), Score 216.0, E value 1.8e-61, N 1
```

```
cDNA EST yk15c4.3 comes from this gene; cDNA EST
EMBL:fo1018 comes from this gene
cDNA EST yk28c9.3 comes from this gene; cDNA EST yk32b2.3
comes from this gene
cDNA EST yk15c4.5 comes from this gene; cDNA EST yk25h10.5
comes from this gene
cDNA EST yk28c9.5 comes from this gene; cDNA EST yk88d7.3
comes from this gene
cDNA EST yk68q2.3 comes from this gene; cDNA EST yk87o12.5
comes from this gene
cDNA EST yk88d7.5 comes from this gene; cDNA EST yk68q2.5
comes from this gene
  cDNA EST yk32c2.3 comes from this gene; cDNA EST
yk161e2.5 comes from this gene
cDNA EST yk169f8.3 comes from this gene; cDNA EST
yk169f8.5 comes from this gene
cDNA EST yk449h9.3 comes from this gene; cDNA EST
yk449h9.5 comes from this gene
cDNA EST yk431f12.3 comes from this gene; cDNA EST
yk431f12.5 comes from this gene
cDNA EST yk419h3.3 comes from this gene; cDNA EST
yk419h3.5 comes from this gene
cDNA EST yk348h7.3 comes from this gene; cDNA EST
yk348h7.5 comes from this gene
cDNA EST yk348e7.3 comes from this gene; cDNA EST
yk348e7.5 comes from this gene
cDNA EST yk324a11.3 comes from this gene; cDNA EST
yk324a11.5 comes from this gene
cDNA EST yk309a2.3 comes from this gene; cDNA EST
yk309a2.5 comes from this gene
cDNA EST yk323b11.3 comes from this gene; cDNA EST
yk323b11.5 comes from this gene
cDNA EST yk315d9.3 comes from this gene; cDNA EST
yk315d9.5 comes from this gene
cDNA EST yk288a8.3 comes from this gene; cDNA EST
yk288a8.5 comes from this gene
cDNA EST yk212d9.3 comes from this gene; cDNA EST
yk220a10.3 comes from this gene
cDNA EST yk220a10.5 comes from this gene; cDNA EST
yk342q2.3 comes from this gene
cDNA EST yk214b1.5 comes from this gene; cDNA EST
yk226t11.5 comes from this gene
cDNA EST yk345a8.5 comes from this gene; cDNA EST
yk405f7.5 comes from this gene
cDNA EST yk405h3.5 comes from this gene; cDNA EST
yk447d12.5 comes from this gene
cDNA EST yk552h2.3 comes from this gene; cDNA EST
yk574a5.3 comes from this gene
cDNA EST yk552h2.5 comes from this gene; cDNA EST
yk571q6.5 comes from this gene
cDNA EST yk574a5.5 comes from this gene"
/codon_start 1
/protein_id "CAA92591.1"
/db_xref "GI:3874925"
/db_xref "SPTREMBL:Q18657"
/translation "MIVLCNKCKNISKMSCTNGGRIFAFLSSVSSIIASFIGNFSIR
RRLVNRFEKVPEIHPAQENFTMGSHWISEQERSRGLEAVQQDWRRFIRDMKWISKKRDD
FLISSRSIVNISNALENSIESLNNVSSIINSSPESLSSAALSIASAFIIFFLSNVAIN
RPRLNRRVSRFIPQAIVIGCINGSLSPFLISPTSAAAATFLLSGKAGPMSPSIGSFINV
AAIVLQNAVSSPQHSIPDRSRLNKIPPNISLASSSSPSDAANNDKPIQQRHSILSNVR
LLLAMVRDGFRFLLGDLMLEMVGLLLEAREFQEFFLLLEEAAERLDVEIEGFLAIERP
IDRSFNCMELLELQRLSFKIVFMCLFILIGVAVAWCFSQFSFLNFTFRQRFREEA
EMLKDLQRDNIVRFYDYWESADLCGKFKYIVLVIELMTSGTLKMYLKRCKRINIKVVL
KSWCRQILKGLSFLHTRNPPVIHRDLKCDNIFITGTTGSVKIGDFGLATLKNESERS
VIGTPEFMAPEMYEEMYDESVDVYAFGMCLLEMVTGEYPYSECMNPATIYRKVISGVK
PESFSKIPAQYPETRELIDRCIKVRKEERSTVKQLLVHLFTPEDLLCTRVETLKNKDA
SLNPLNVETIQMGLPVYPLFEEPQYFFLNPSFQFAFDTFNDCPPPVVQQMIPQQHIPD
EPTRMETKQLEQRVDAFRRQMQHRGLLEIKRAEEEEERIREEAEIFEEIPLKAEAEEFE
KEELEKERLEKKAAAAAAANPNIPTPIPTPATPHSSAQQQPTPPPLSTQTSAEIQQSA
QQQSVFVTPMIANTPAMSPTSAQTQPVFSFTSAAVPVFTPMIHVFRPSETPVQNVATTA
APVAANNVPPSPAPFKTFDTQTPTLAQNIVPPTISTDASGLVINTPASIASPSPAPSA
TDVASTTAPVTPAPTPPTTTTDGCAAAASTTTENKEEKRESNKPKVVMETLGCDESRNF
ALVSCRLDTSHKSVTEQFAPGTDKPCTIATKLEAEDCTLKVHVHTVEAQLGEVIQLIN
```

```
                     SDGKKGVGTKLATVLDFNSTPVPLLTAVHPKSSAA ACNLFELLLR..PLSLKC
                     EPNNVQVTNVPFVSQFSNAFSVQSIPPFSSTLVMSFLNSTLSAFFTLAAAEFCFGV
                     TKSADPIATPISSSISTATVIPIVAATPTNITSEPVIVQPITAQVITHLATPSPVSHS
                     LSSNSSPSATTHSNMSSIQSTISVIGKRFTVQPVSQAESGISSSISTPHPEPTPAITS
                     CPPPVPSVPPVVSNGTLNLEVAPKQTPSATNQNVDTQHSSSTASTATLVSETPATVHV
                     TPISVPAPVQEPLVIDHHSDVLTQLDSELKKVSGVSHSASPSTVVESLISMTPQTIPL
                     ACQTVPASIGQAPAVIAAAHAASLIPNASVPQSPSRIDAETGLAGLHEKLEALKMEQD
                     RREDMGDDATGFTTIDGKDFIPIDTLKGLAPALGKVTHAIGREFIPMPPDHPDLIDAS
                     TQQLISPSNPDVLTTMSSAVEGSASSIMIEDIDASTSAVDASMMNSMPTGAQNSTGGI
                     PAAMTLSMDQECAQSMTSSITRNTTGTKLATFENLETALSSILGTHTRQPNAPSSRDE
                     TTAPMTPSFTNERIGGGGGGGAISFSIGTPPSHSPFPVSECDYDLKGQMDLESEDPEV
                     IQMIVRHRMEQHKLLEKQRVEIFRLRSKIPVPPATSVNPEMICDDEAIVTLTALQSAL
                     GNASLSIPASPPPNTFTTPVNTTVIDSPVIATEMTMPSQFSTKSPNVSVSSEHEDNQSA
                     PPRHHHHQPHPPHHPHLQNHYHPDQNHTSATAPCPSAHVQLQAVSNNNVNPLRGPFHP
                     VSSQIPPQA"
     gene            join(30950..31094,31814..31774,31826..32042,32101..32212,
                     32259..32364,32727..33020,33104..33258,33306..33369)
                     /gene="C46C2.2"
     CDS             join(30950..31094,31814..31774,31826..32042,32101..32212,
                     32259..32364,32727..33020,33104..33258,33306..33369)

alignment_scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_234_242 x CEC46C2   ..

 Align seq 1/1  to: CEC46C2  from: 1  to: 41115

       3 ThrAspSerGluValAspSer 9
         ||||||||||||||||||||
   38400 ACTGACTCGGAAGTGGACACT 38420

seq_name: qb_pr2:AC008530

seq_documentation_block:
LOCUS       AC008530    89095 bp    DNA             PRI       31-MAY-2000
DEFINITION  Homo sapiens chromosome 19 clone CTC-478M6, complete sequence.
ACCESSION   AC008530
VERSION     AC008530.7  GI:8122100
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 89095)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 89095)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 89095)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAY-2000) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On May 31, 2000 this sequence version replaced gi:7711288.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >-40 99.9% of Sequence;
            Estimated Total Number of Errors is 0.1.
FEATURES             Location/Qualifiers
     source          1..89095
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="CTC-478M6"
BASE COUNT    27679 a   18534 c   17203 g   25679 t
ORIGIN


alignment_scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_234_242 x AC022428   ..

 Align seq 1/1  to: AC008530  from: 1  to: 89095

       3 ThrAspSerGluValAspSer 9
         ||||||||||||||||||||
   84199 ACAGACTCTGAGGTCGACTCA 84219

seq_name: qb_pr1:AC022428

seq_documentation_block:
LOCUS       AC022428    94890 bp    DNA             PRI       01-JUL-2000
DEFINITION  Homo sapiens chromosome 5 clone CTD-2372H16, complete sequence.
ACCESSION   AC022428
VERSION     AC022428.6  GI:8886983
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 94890)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 94890)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03 FEB 2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 94890)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUL-2000) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jul 1, 2000 this sequence version replaced gi:7711744.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >-40 99.8% of Sequence;
            Estimated Total Number of Errors is 0.4.
FEATURES             Location/Qualifiers
     source          1..94890
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2372H16"
BASE COUNT    27974 a   17653 c   19522 g   29741 t
ORIGIN


alignment_scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_234_242 x AC022428/rev   ..

 Align seq 1/1  to reverse of: AC022428  from: 1  to: 94890

       1 SerIrpIhrAspSerGluVal 7
```

```
          !!!!!!!!!!!!!!!!!!!!
    44467 AGCTGCACTCACACTGAAGTA 44447

seq name: qb_htql:AC008879

seq documentation block:
LOCUS       AC008879    124682 bp    DNA              HTG       18 JUL 2000
DEFINITION  Homo sapiens chromosome 5 clone CTD 2210C1, WORKING DRAFT SEQUENCE,
            6 ordered pieces.
ACCESSION   AC008879
VERSION     AC008879.4  GI:9256066
KEYWORDS    HTG; HTGS PHASE2; HTGS DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 124682)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 124682)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG 1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jul 18, 2000 this sequence version replaced gi:7209472.
              -      Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: https://www.jgi.doe.gov
            ----
            Project Information
            Center Project Name: 708248
            Center clone name: CTTB HI 2210C1

            Summary Statistics
            Consensus quality: 120664 bases at least Q40
            Consensus quality: 123440 bases at least Q30
            Consensus quality: 123897 bases at least Q20
            Estimated insert size: 127000; pulse field gel estimation
            Estimated insert size: 124482; sum of contigs estimation
            Quality coverage:  6.35 in Q20 bases; pulse field gel estimation
            Quality coverage:  6.48 in Q20 bases; sum of contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submitter
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1     22453: contig of 22453 bp in length
            *    22454     22553: gap of unknown length
            *    22554     24180: contig of 1627 bp in length
            *    24181     24280: gap of unknown length
            *    24281     94694: contig of 70414 bp in length
            *    94695     94794: gap of unknown length
            *    94795    107507: contig of 12713 bp in length
            *   107508    107607: gap of unknown length
            *   107608    122781: contig of 15174 bp in length
            *   122782    122881: gap of unknown length
            *   122882    124682: contig of 1801 bp in length.
FEATURES             Location/Qualifiers
     source          1..124682
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD 2210C1"
                     /clone_lib="CalTech human BAC library D"
BASE COUNT    45730 a  24040 c  25229 g  40182 t    501 others
ORIGIN
```

```
alignment scores:
            Quality:     7.00              Length:       7
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 QS 09 126-945B-2 COPY 234_242 x AC008879/rev   ..

 Align seq 1/1:  to reverse of: AC008879  from: 1  to: 124682

        1 SerIleThrAspSerGluVal 7
          !!!!!!!!!!!!!!!!!!!!
    39430 AGCTGCACTCACACTGAAGTA 39410

seq name: qb_pr3:AC016591

seq documentation block:
LOCUS       AC016591   147362 bp    DNA              PRI       30 MAR 2001
DEFINITION  Homo sapiens chromosome 5 clone CTD 2303D5, complete sequence.
ACCESSION   AC016591
VERSION     AC016591.5  GI:13489151
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 147362)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 147362)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 147362)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Mar 30, 2001 this sequence version replaced gi:7710213.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www.shgc.stanford.edu
            Quality: Phrap Quality > 40 99.9% of Sequence;
            Estimated Total Number of Errors is 0.1.
FEATURES             Location/Qualifiers
     source          1..147362
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD 2303D5"
BASE COUNT    43808 a  27622 c  25701 g  40231 t
ORIGIN


alignment scores:
            Quality:     7.00              Length:       7
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 QS 09 126-945B-2 COPY 234_242 x AC016591   ..

 Align seq 1/1:  to: AC016591  from: 1  to: 147362

        1 SerIleThrAspSerGluVal 7
          !!!!!!!!!!!!!!!!!!!!
    17718 AGCTGCACTCACACTGAAGTA 17738

seq name: qb_htq5:AC016214
```

```
seq_documentation_block:
LOCUS       AC016212   160217 bp    DNA              HTG       16-MAR-2000
DEFINITION  Homo sapiens clone RP11-25P7, WORKING DRAFT SEQUENCE, 9 unordered
            pieces.
ACCESSION   AC016212
VERSION     AC016212.4  GI:7249075
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 160217)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-25P7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 160217)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Bogoslavsky,L., Bouhalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Dzamba,M., Fenelon,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Hoaford,A., Horton,L.,
            Howland,J.C., Johnson,E., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Liew,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGark,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,I., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 16, 2000 this sequence version replaced gi:6649341.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            --------- ---- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www.seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            ------------- Project Information
                Center project name: L4619
                Center clone name: 25_P_7
            ------ ------ Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 155915 bases at least Q40
                Consensus quality: 158008 bases at least Q30
                Consensus quality: 158845 bases at least Q20
                Insert size: 158000; agarose-tp
                Insert size: 159417; sum of contigs
                Quality coverage: 5.1 in Q20 bases; agarose tp
                Quality coverage: 5.1 in Q20 bases; sum of contigs
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1800: contig of 1800 bp in length
            *     1801  1900: gap of      100 bp
            *     1901     4600: contig of 2700 bp in length
            *     4601  4700: gap of      100 bp
            *     4701     8787: contig of 4087 bp in length
            *     8788  8887: gap of      100 bp
            *     8888    16192: contig of 7305 bp in length
            *    16193  16292: gap of      100 bp
            *    16293    29642: contig of 13350 bp in length
            *    29643  29742: gap of      100 bp
            *    29743    58221: contig of 28479 bp in length
            *    58222  58321: gap of      100 bp
            *    58322    75019: contig of 21698 bp in length
            *    75020  75119: gap of      100 bp
            *    75120    99739: contig of 24620 bp in length
            *    99740  99839: gap of      100 bp
            *    99840   160217: contig of 60378 bp in length.
FEATURES             Location/Qualifiers
     source          1..160217
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-25P7"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..1800
                     /note="assembly_fragment"
     misc_feature    1901..4600
                     /note="assembly_fragment"
     misc_feature    4701..8787
                     /note="assembly_fragment"
     misc_feature    8888..16192
                     /note="assembly_fragment
                     clone_end:17
                     vector_side:right"
     misc_feature    16293..29642
                     /note="assembly_fragment"
     misc_feature    29743..58221
                     /note="assembly_fragment"
     misc_feature    58322..75019
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    75120..99739
                     /note="assembly_fragment"
     misc_feature    99840..160217
                     /note="assembly_fragment"
BASE COUNT    45570 a   33216 c   33535 g   47092 t     804 others
ORIGIN


alignment_scores:
              Quality:    7.00              Length:       7
                Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 ns-ow-126-945b-2_COPY_234_242 x AC016212/rev   ..

 Align seq 1/1  to reverse of: AC016212  from: 1  to: 160217

        3 ThrAspSerGluValAspSer 9
          |||| ||||| |||||||||
   148865 ACAGATCTGAAGTGGACAGT 148845

seq_name: gb_pri8:HS125H2

seq_documentation_block:
LOCUS       HS125H2     173513 bp    DNA              PRI       12-DEC-1999
DEFINITION  Human DNA sequence from clone CTA-125H2 on chromosome 22q11-12.
            Contains part of the gene for a novel protein similar to KIAA0216
            and myosin heavy chain, ESTs, GSSs and a CpG island, complete
            sequence.
ACCESSION   298949
VERSION     298949.1  GI:1508990
KEYWORDS    HTG; CpG island; KIAA0216; myosin.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 173513)
  AUTHORS   Lloyd,D.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On May 30, 1998 this sequence version replaced gi:2598540.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 22, constructed by the Sanger Centre Chromosome 22
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr22.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; SW:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep.shtml.
            from the human BAC library described in D.J. Kim et al. (1996)
            Genomics 34, 213-218.
            VECTOR: pBAC108L
            This sequence is the entire insert of clone CTA-125H2. The start of
            this sequence overlaps with sequence AL022329. The end of this
            sequence overlaps with sequence AL000000.
FEATURES                Location/Qualifiers
     source             1..173513
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="22"
                        /map="q11-12"
                        /clone="CTA-125H2"
                        /clone_lib="CITY8SK.A1"
     repeat_region      231..371
                        /note="MIR repeat: matches 1..142 of consensus"
     gene               join(564..710,9002..9161,11467..11550,12697..12919,
                        16737..16846,33587..33652,39953..39981,41456..41599,
                        43226..43315,56023..56142,56571..56679,60423..60547,
                        63574..63728,67880..68004,68898..69098,73589..73699,
                        76213..76317,86524..86676,111403..111516,112978..113094,
                        115633..115716,117513..117689,120445..120640,
                        157629..157759)
                        /gene="bK125H2.1"
     CDS                join(564..710,9002..9161,11467..11550,12697..12919,
                        16737..16846,33587..33652,39953..39981,41456..41599,
                        43226..43315,56023..56142,56571..56679,60423..60547,
                        63574..63728,67880..68004,68898..69098,73589..73699,
                        76213..76317,86524..86676,111403..111516,112978..113094,
                        115633..115716,117513..117689,120445..120640,
                        157629..157759)
                        /gene="bK125H2.1"
                        /note="match: cDNAs: Em:D86970
                        match: ESTs: Em:ZI8346 Em:AA504934 Em:R41006
                        match: proteins: Tr:O60772 Tr:Q92614
                        match: genomic DNA: Em:H55072 Em:H55041 Em:H55303"
                        /codon_start=1
                        /evidence=NOT_experimental
                        /product="bK125H2.1 (novel protein similar to KIAA0216 and
                        myosin heavy chain)"
                        /protein_id="CAB11606.2"
                        /db_xref="GI:6572192"
                        /db_xref="SPTREMBL:O60772"
                        /translation="VRTPAGSCAQDAWCTFWVTHFFVHVFCSVSVVTFFTCAAFFFF
                        GAGTEGSSALRTCFQPLQCETFHQLGWDPVRYDLTGWLHRAKPNLSALDAPQVLHQSK
                        REELRSLFQARAKLPPVCRAVACLEGTSQQALQRSRMVRFTFASSLAAVRKKAPCSQT
                        KLQMDALTSMIKRSRLHFTHCTVPNPVVESRSGQESPFPPQPGRDKPGAGGPLALDTP
```

```
                        ALRVQLAGFHILEALRLHRTGYADHMGLTKFRRQFQVLLAPLLKKELMSTSECTEERKA
                        VELLLELLDLEKKAVAVGHSQVELKAGVTSRELKQFKKLVSQSTVLFQAACKGFLSRQ
                        EFKELETRKLAAQCTQKNVAVFLAVFDWPWWQLLGSTQPLLSALTCTEQLRAKEEELL
                        LLERSLEKSEKLRNELRQNNIQLLESKIAQLISDLADERFEGDVACTQLESERAERLQA
                        TFFVQELKASHLEVQKKSIEVNKQLLEAQCKTQLNGLLENCTGGAISEWQMKFDCAQME
                        NLFLKKKLQQVTEKFLISELIAKKELEQKLQSLQSAYICAKAMAHQLERKCHHLICDLE
                        DICVLLENQQSRNHELEAKQKKFDLQLAQAIQGSVEEKGLKEKVTQENTSVKWELGQL
                        VQQLEQKFQEASQLKQQVEMLQHKREELQSPSDGENCVAGLEEELWELESSALEQQK
                        TQSQQENITKQLPQLEQRFEELETERMKQMHQKDREDQEFFELEDVRQSCQKRLHQLEMQ
                        LEQEYEEKQMVLHEKQDLEGLIGTLCQVIGHRGDIDVEKRLKKDLRRTHALLSDVQLLL
                        CIMEIKRISVSKEELERVHSQFQSFAKCPEALKIQKVLIAQLESMDSELENMIKNES
                        LVDEQLYKLQFEEKAQLEKRIHEDQQDELNELMQEHEDLIAQSAADIDCQIQELQLEEA
                        KEEEKEIQPQCQQMSNKLLSYLQLQVAQMKTEYIEQSIVDKAIVSPQFAVIVDLFNKI
                        EDQKVQCKKEFVIVIPLPDSLIKMCFELSQAAISFSQQKFSSQYYQRKIFELKADMEE
                        VQHAASGCMLLKYVLDAAVEQLQGIEHAPLQAALDVASDSDIE
                        "
repeat_region       855..924
                    /note="L2 repeat: matches 2086..2749 of consensus"
repeat_region       1660..1713
                    /note="18 copies 3-mer tqa 72 conserved"
repeat_region       1946..2233
                    /note="AluSq repeat: matches 1..287 of consensus"
repeat_region       2234..2278
                    /note="19 copies 3-mer taa 100 conserved"
repeat_region       2362..2423
                    /note="L2 repeat: matches 2671..2734 of consensus"
repeat_region       2489..2836
                    /note="L2 repeat: matches 1257..1637 of consensus"
repeat_region       3130..3344
                    /note="L1PA11 repeat: matches 5960..6169 of consensus"
repeat_region       3384..3475
                    /note="23 copies 4-mer atat 72 conserved"
repeat_region       3392..3473
                    /note="11 copies 2-mer at 76 conserved"
repeat_region       3483..3743
                    /note="AluSq repeat: matches 21..283 of consensus"
repeat_region       3750..4067
                    /note="MER7A repeat: matches 1..329 of consensus"
repeat_region       4073..4252
                    /note="45 copies 4-mer tgtg 73 conserved"
repeat_region       4077..4250
                    /note="87 copies 2-mer tg 73 conserved"
repeat_region       4253..4552
                    /note="75 copies 4-mer gtat 75 conserved"
repeat_region       4555..4640
                    /note="MER5L repeat: matches 189..274 of consensus"
repeat_region       5228..5472
                    /note="L2 repeat: matches 2434..2694 of consensus"
repeat_region       5497..5647
                    /note="L2 repeat: matches 2286..2442 of consensus"
repeat_region       7015..7116
                    /note="L2 repeat: matches 2561..2669 of consensus"
repeat_region       7200..7239
                    /note="L2 repeat: matches 2580..2619 of consensus"
repeat_region       7658..7961
                    /note="AluSq repeat: matches 1..305 of consensus"
repeat_region       8081..8163
                    /note="L2 repeat: matches 2619..2713 of consensus"
repeat_region       8230..8419
                    /note="MIR repeat: matches 57..245 of consensus"
misc_feature        complement(9497..9973)
                    /note="match: GSS: Em:AQ200189"
repeat_region       9769..9950
                    /note="MIR repeat: matches 54..250 of consensus"
repeat_region       10305..10353
                    /note="MIR repeat: matches 92..149 of consensus"
repeat_region       10504..10811
                    /note="AluSc repeat: matches 1..309 of consensus"
repeat_region       11638..11776
                    /note="MIR repeat: matches 34..188 of consensus"
repeat_region       12067..12375
                    /note="AluSc repeat: matches 1..309 of consensus"
repeat_region       12426..12515
```

```
                     /note="MIR repeat: matches 47, 136 of consensus"
     repeat_region   13064, .13180
                     /note "MIR repeat, matches 28, 134 of consensus"
     misc_feature    complement(13160, .13420)
                     /note="match: GSS: Em:AQ009682"
     repeat_region   13204   13243
                     /note "10 copies 4 mer acac 95 conserved"
     repeat_region   13246, .13314
                     /note="L1PA7 repeat: matches 6057, .6125 of consensus"
     misc_feature    13444, .13868
                     /note="match: GSS: Em:AQ600862"
     misc_feature    13447, .13881
                     /note "match: GSS: Em:AQ409072"
     misc_feature    13449, .13908
                     /note="match: GSS: Em:AQ819687"
     repeat_region   13535, .13778
                     /note "L2 repeat, matches 2271, .2554 of consensus"
     repeat_region   14793, .14826
                     /note="L2 repeat: matches 2695, .2729 of consensus"
     repeat_region   14901, .15097
                     /note "MIR repeat: matches 4, .221 of consensus"
     repeat_region   15155, .15308
                     /note="MIR repeat, matches 31, .190 of consensus"
     repeat_region   15636, .15841
                     /note="MIR repeat: matches 44, .262 of consensus"
     repeat_region   15848   16027
                     /note="AluJb repeat: matches 112, .297 of consensus"
     repeat_region   16056, .16185
                     /note="L2 repeat: matches 2455, .2487 of consensus"
     repeat_region   16918, .17021
                     /note="AluJ/FLAM repeat, matches 2, .101 of consensus"
     repeat_region   17444, .17555
                     /note="L2 repeat: matches 2363, .2475 of consensus"
     repeat_region   17623, .17669
                     /note "L2 repeat: matches 2665, .2711 of consensus"
     repeat_region   17860, .17971
                     /note="MIR repeat: matches 77, .192 of consensus"
     repeat_region   18033, .18154
                     /note "MIR repeat, matches 9, .146 of consensus"
     misc_feature    18090, .18674

alignment_scores:
             Quality:    7.00              Length:       7
               Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_234_242 x HS125B2/rev

 Align seq 1/1  to reverse of: HS125B2  from: 1  to: 179104

       1 SerIrpThrAspSerGluVal 7
         |||||||||||||||||||||
   10856 TCATGGACAGATTCTCAGGTT 10836

seq_name: gb_htg20:AL158154

seq_documentation_block:
LOCUS       AL158154    179104 bp    DNA             HTG       08 APR 2001
DEFINITION  Homo sapiens chromosome 9 clone RP11 483M4, *** SEQUENCING IN
            PROGRESS ***, 3 unordered pieces.
ACCESSION   AL158154
VERSION     AL158154.22  GI:13567910
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN, HTGS_DRAFT; HTGS FULLTEXT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 179104)
  AUTHORS   Clark,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07 APR 2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 9, 2001 this sequence version replaced gi:13373889.
                ~~~~~~~~~~ Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
              ~~~~~~~~~~ Project Information
            Center project name: DA383M4
              ~~~~~~~~~~ Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: dye terminator BigDye; 100% of reads
            Consensus quality: 178216 bases at least Q40
            Consensus quality: 178510 bases at least Q30
            Consensus quality: 178713 bases at least Q20
            Insert size: 176604, sum of contigs
            Insert size: 180789; 12.4% error; agarose fp
            Quality coverage: 10.6% in Q20 bases; sum of contigs quality
            coverage: 10.56x in Q20 bases; agarose fp
              ~~~~~~~~~~~~~~
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    92910: contig of 92910 bp in length
            *   92911 93010: gap of      100 bp
            *   93011    97935: contig of 4925 bp in length
            *   97936 98035: gap of      100 bp
            *   98036   179104: contig of 81069 bp in length.
FEATURES             Location/Qualifiers
     source          1..179104
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-483M4"
                     /clone_lib="RPCI-11.2"
     misc_feature    1..92910
                     /note="assembly_fragment:02755"
     misc_feature    93011..97935
                     /note="assembly_fragment:03138"
     misc_feature    98036..179104
                     /note="assembly_fragment:04341"
BASE COUNT    [illegible]  48255 g  52318 t    200 others
ORIGIN


alignment_scores:
             Quality:    7.00              Length:       7
               Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_234_242 x AL158154/rev

 Align seq 1/1  to reverse of: AL158154  from: 1  to: 179104

       3 ThrAspSerGluValAspSer 9
         |||||||||||||||||||||
  157648 ACAGACTCAGAGGTCGACAGT 157628

seq_name: gb_htg16:AC079219

seq_documentation_block:
LOCUS       AC079219    181008 bp    DNA             HTG       24 AUG 2000
DEFINITION  Mus musculus chromosome 5 clone RP23-47M20 strain C57BL/6J,
            WORKING DRAFT SEQUENCE, 19 unordered pieces.
ACCESSION   AC079219
VERSION     AC079219.1  GI:9887659
```

```
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 181008)
  AUTHORS   Bockstroce-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
            Bouffard,G.G., Dietrich,N.L., Eagle,W.C., Gupta,J., Ho,S. L.,
            Idol,J., Lee-Lin,S. Q., Lequaspi,R., Lim,M., Maduro,Q.L.,
            Maduro,V.B., Mastrian,S.D., McCloskey,J.C., Medowell,L.,
            Pearson,R., Stantripop,S., Summers,T.L., Thomas,L.W., Thomas,P.J.,
            Tiongson,E.E., Touchman,J.W., Tran,J.L., Vogt,J.L., Walker,M.A.,
            Wetherby,K.D. and Green,E.D.
  TITLE     NISC Mouse Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 181008)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (24 AUG 2000) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT                 Genome Center
            Center: NIH Intramural Sequencing Center
            Center code: NISC
            Web site: http://www.nisc.nih.gov
            Contact: nisc-mouse@nhgri.nih.gov
                    Project Information
            Center project name: tm
            Center clone name: 454D20
                    Summary Statistics
            Sequencing vector: plasmid; n/a; 100% of reads
            Chemistry: Dye terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 169848 bases at least Q40
            Consensus quality: 172456 bases at least Q30
            Consensus quality: 173984 bases at least Q20
            Insert size: 219000; agarose fp
            Insert size: 186000; pulse field gel
            Insert size: 179208; sum of contigs
            Quality coverage: 4.04x in Q20 bases; agarose fp
            Quality coverage: 4.56x in Q20 bases; pulse field gel
            Quality coverage: 4.74x in Q20 bases; sum of contigs
                    .
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 19 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1      3541: contig of 3541 bp in length
            *    3542      3641: gap of unknown length
            *    3642      8844: contig of 5203 bp in length
            *    8845      8944: gap of unknown length
            *    8945     14381: contig of 5437 bp in length
            *   14382     14481: gap of unknown length
            *   14482     19046: contig of 4565 bp in length
            *   19047     19146: gap of unknown length
            *   19147     23850: contig of 4704 bp in length
            *   23851     23950: gap of unknown length
            *   23951     28382: contig of 4432 bp in length
            *   28383     28482: gap of unknown length
            *   28483     35768: contig of 7286 bp in length
            *   35769     35868: gap of unknown length
            *   35869     41476: contig of 5608 bp in length
            *   41477     41576: gap of unknown length
            *   41577     47592: contig of 6016 bp in length
            *   47593     47692: gap of unknown length
            *   47693     53389: contig of 5697 bp in length
            *   53390     53489: gap of unknown length
            *   53490     63729: contig of 10240 bp in length
            *   63730     63829: gap of unknown length
            *   63830     74096: contig of 10267 bp in length
            *   74097     74196: gap of unknown length
            *   74197     89485: contig of 15289 bp in length
            *   89486     89585: gap of unknown length
            *   89586    101436: contig of 11751 bp in length
            *  101437    101446: gap of unknown length
            *  101447    111623: contig of 10187 bp in length
            *  111624    111723: gap of unknown length
            *  111724    127083: contig of 15360 bp in length
            *  127084    127183: gap of unknown length
            *  127184    140040: contig of 12857 bp in length
            *  140041    140140: gap of unknown length
            *  140141    158082: contig of 17942 bp in length
            *  158083    158182: gap of unknown length
            *  158183    181008: contig of 22826 bp in length.
FEATURES             Location/Qualifiers
     source          1..181008
                     /organism="Mus musculus"
                     /strain="C57BL6/J"
                     /db_xref="taxon:10090"
                     /chromosome="5"
                     /clone="RP23-454D20"
                     /clone_lib="RPCI mouse BAC library 23"
     misc_feature    1..3541
                     /note="assembly_fragment"
     misc_feature    3642..8844
                     /note="assembly_fragment"
     misc_feature    8945..14381
                     /note="assembly_fragment"
     misc_feature    14482..19046
                     /note="assembly_fragment"
     misc_feature    19147..23850
                     /note="assembly_fragment"
     misc_feature    23951..28382
                     /note="assembly_fragment"
     misc_feature    28483..35768
                     /note="assembly_fragment"
     misc_feature    35869..41476
                     /note="assembly_fragment"
     misc_feature    41577..47592
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    47693..53389
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    53490..63729
                     /note="assembly_fragment"
     misc_feature    63830..74096
                     /note="assembly_fragment"
     misc_feature    74197..89485
                     /note="assembly_fragment"
     misc_feature    89586..101436
                     /note="assembly_fragment"
     misc_feature    101447..111623
                     /note="assembly_fragment"
     misc_feature    111724..127083
                     /note="assembly_fragment"
     misc_feature    127184..140040
                     /note="assembly_fragment"
     misc_feature    140141..158082
                     /note="assembly_fragment"
     misc_feature    158183..181008
                     /note="assembly_fragment"
BASE COUNT    48542 a   40713 c   39639 g   50289 t    1825 others
ORIGIN


alignment scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000
```

```
alignment_block:
 US-09-126-945B-2_COPY_234_242 x AC079219  ..

 Align seq 1/1  to: AC079219  from: 1  to: 181008

        3 ThrAspSerGluValAspSer 9
          ||||||||||||||||||||||
   156397 ACTGACAGTGAAGTAGATAGT 156417

seq_name: qb_htq5:AC016583

seq_documentation_block:
LOCUS       AC016583    182082 bp    DNA             HTG       05-MAY-2000
DEFINITION  Homo sapiens chromosome 5 clone CITB-251A1, WORKING DRAFT SEQUENCE,
            43 unordered pieces.
ACCESSION   AC016583
VERSION     AC016583.3  GI:7710215
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 182082)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 182082)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     on May 5, 2000 this sequence version replaced gi:7341912
            ------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 824926
            Center clone name: CITB-E1_251A1
            ------
            Summary Statistics
            Consensus quality: 139424 bases at least Q40
            Consensus quality: 161110 bases at least Q30
            Consensus quality: 166832 bases at least Q20
            Estimated insert size: 188000; pulse field gel estimation
            Estimated insert size: 177882; sum-of-contigs estimation
            Quality coverage:  8.48 in Q20 bases; pulse field gel estimation
            Quality coverage:  8.68 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 43 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2159: contig of 2159 bp in length
            *     2160     2259: gap of unknown length
            *     2260     5159: contig of 2900 bp in length
            *     5160     5259: gap of unknown length
            *     5260     7472: contig of 2213 bp in length
            *     7473     7572: gap of unknown length
            *     7573     9890: contig of 2318 bp in length
            *     9891     9990: gap of unknown length
            *     9991    12143: contig of 2153 bp in length
            *    12144    12243: gap of unknown length
            *    12244    14965: contig of 2722 bp in length
            *    14966    15065: gap of unknown length
            *    15066    17080: contig of 2015 bp in length
            *    17081    17180: gap of unknown length
            *    17181    19256: contig of 2076 bp in length
            *    19257    19356: gap of unknown length
            *    19357    21458: contig of 2102 bp in length
            *    21459    21558: gap of unknown length
            *    21559    24820: contig of 3262 bp in length
            *    24821    24920: gap of unknown length
            *    24921    27117: contig of 2197 bp in length
            *    27118    27217: gap of unknown length
            *    27218    29459: contig of 2242 bp in length
            *    29460    29459: gap of unknown length
            *    29460    32927: contig of 3468 bp in length
            *    32928    33027: gap of unknown length
            *    33028    35059: contig of 2032 bp in length
            *    35060    35159: gap of unknown length
            *    35160    38509: contig of 3350 bp in length
            *    38510    38609: gap of unknown length
            *    38610    41536: contig of 2927 bp in length
            *    41537    41636: gap of unknown length
            *    41637    44090: contig of 2454 bp in length
            *    44091    44190: gap of unknown length
            *    44191    46766: contig of 2576 bp in length
            *    46767    46866: gap of unknown length
            *    46867    49006: contig of 2140 bp in length
            *    49007    49106: gap of unknown length
            *    49107    51958: contig of 2852 bp in length
            *    51959    52058: gap of unknown length
            *    52059    54833: contig of 2775 bp in length
            *    54834    54933: gap of unknown length
            *    54934    59266: contig of 4333 bp in length
            *    59267    59366: gap of unknown length
            *    59367    62275: contig of 2909 bp in length
            *    62276    62375: gap of unknown length
            *    62476    65697: contig of 3322 bp in length
            *    65698    65797: gap of unknown length
            *    65798    69858: contig of 4061 bp in length
            *    69859    69958: gap of unknown length
            *    69959    74075: contig of 4117 bp in length
            *    74076    74175: gap of unknown length
            *    74176    79360: contig of 5185 bp in length
            *    79361    79460: gap of unknown length
            *    79461    83393: contig of 3933 bp in length
            *    83394    83493: gap of unknown length
            *    83494    88250: contig of 4757 bp in length
            *    88251    88350: gap of unknown length
            *    88451    92740: contig of 4390 bp in length
            *    92741    92840: gap of unknown length
            *    92841    97160: contig of 4320 bp in length
            *    97161    97260: gap of unknown length
            *    97261   101740: contig of 4480 bp in length
            *   101741   101840: gap of unknown length
            *   101841   105428: contig of 3588 bp in length
            *   105429   105528: gap of unknown length
            *   105529   110875: contig of 5347 bp in length
            *   110876   110975: gap of unknown length
            *   110976   115266: contig of 4291 bp in length
            *   115267   115366: gap of unknown length
            *   115367   120061: contig of 4695 bp in length
            *   120062   120161: gap of unknown length
            *   120162   127060: contig of 6899 bp in length
            *   127061   127160: gap of unknown length
            *   127161   133003: contig of 5843 bp in length
            *   133004   133103: gap of unknown length
            *   133104   138779: contig of 5676 bp in length
            *   138780   138879: gap of unknown length
            *   138880   147356: contig of 8477 bp in length
            *   147357   147456: gap of unknown length
            *   147457   156954: contig of 9498 bp in length
            *   156955   157054: gap of unknown length
            *   157055   167993: contig of 10939 bp in length
            *   167994   168093: gap of unknown length
            *   168094   182082: contig of 13989 bp in length.
FEATURES             Location/Qualifiers
     source          1..182082
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

```
                     /chromosome "5"
                     /clone "CTD-2501A1"
                     /clone_lib "Caltech human BAC library D"
BASE COUNT     52170 a   46139 c   46457 g   53084 t    4232 others
ORIGIN


alignment scores:
            Quality:     7.00              Length:        7
              Ratio:    1.000                Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-126-945B-2-COPY_234_242 x AC016583/rev

 Align seq 1/1  to reverse of: AC016583  from: 1  to: 182082

        1 SerTrpThrAspSerGluVal 7
          |||||||||||||||||||||
   114815 AGCTGGACTGACAGTGAAGTA 114815
```

```
OM of: US-09-126-945B-2_COPY_234_242.rng [ ] Y Score [ ] Length [ ]

Date: Sep 11, 2001 9:49 PM

About: Results were produced by the SeqWare Search version 1.5
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
  -MODEL=trembl_p2n.model -DEV xlp
  -Q=/cgn2_1/USPTO_spool/US09126945_jma_20010911_132511_1/5_app_query_list_1.247
  -DB=N_Geneseq_0601 -QFMT=fastap SUFFIX=olip2n.rng GAPOP 4.500
  -GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
  -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
  -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
  -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
  -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality
  -THR_MIN=1 -ALIGN=15 MODE=LOCAL OUTFMT=pir NORM=ext -MINLEN=0
  -MAXLEN=2000000000 -USER=US09126945_JMA_1_308 NCPU=6 -ICPU=3
  -LONGLOC NO_XLIPXY WAIT THREADS=1

Search information block:
Query: US-09-126-945B-2_COPY_234_242
Query length: 9
Database: N_Geneseq_0601:*
Database sequences: 730101
Database length: 313950809
Search time (sec): 680.150000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT 60.000


score_list:
Sequence              Strd Orig    ZScore   EScore Len  ! Documentation  ..
/SIDS1/geodata/geneseq/geneseqn/NA2030.DAT AAA06617 +     9.00   100.66     0.0110     278  !
/SIDS1/geodata/geneseq/geneseqn/NA2030.DAT AAE21878 +     8.00   181.21     0.0365    1087  !
/SIDS1/geodata/geneseq/geneseqn/NA2000.DAT AAA75641 +     8.00   177.54     0.0413    1064  !
/SIDS1/geodata/geneseq/geneseqn/NA2001.DAT AAQ84251 +     8.00   177.54     0.0413    1064  !
/SIDS1/geodata/geneseq/geneseqn/NA2001.DAT AAQ84226 +     8.00   173.70     0.0529    1017  !
/SIDS1/geodata/geneseq/geneseqn/NA1995.DAT AAT24181 +     6.00   127.19     99.46     396  !
/SIDS1/geodata/geneseq/geneseqn/NA2000.DAT AAT10170 +     6.00   126.97     42.16     426  !
/SIDS1/geodata/geneseq/geneseqn/NA2002.DAT AAT79236 +     6.00   123.63     64.39     697  !
/SIDS1/geodata/geneseq/geneseqn/NA2000.DAT AAP07726 +     6.00   123.41     66.55     716  !
/SIDS1/geodata/geneseq/geneseqn/NA2001.DAT AAP54219 +     6.00   121.07     99.80    1007  !
/SIDS1/geodata/geneseq/geneseqn/NA1988.DAT AAN92450 +     6.00   119.67    107.54    1236  !
/SIDS1/geodata/geneseq/geneseqn/NA2000.DAT AAP14511 +     6.00   117.83    126.19    1617  !
/SIDS1/geodata/geneseq/geneseqn/NA1999.DAT AAX88027 +     6.00   116.22    167.27    2043  !
/SIDS1/geodata/geneseq/geneseqn/NA1987.DAT AAN70807 +     6.00   115.01    195.53    2348  !
/SIDS1/geodata/geneseq/geneseqn/NA1993.DAT AAQ37484 +     6.00   115.01    195.53    2348  !
/SIDS1/geodata/geneseq/geneseqn/NA1996.DAT AAX89075 +     6.00   114.77    201.57    2526  !
/SIDS1/geodata/geneseq/geneseqn/NA1996.DAT AAX89074 +     6.00   114.77    201.57    2526  !
/SIDS1/geodata/geneseq/geneseqn/NA1987.DAT AAN70805 +     6.00   114.61    204.91    2574  !
/SIDS1/geodata/geneseq/geneseqn/NA1987.DAT AAN70804 +     6.00   114.61    205.61    2575  !
/SIDS1/geodata/geneseq/geneseqn/NA1986.DAT AAN70434 +     6.00   114.61    205.15    2577  !
/SIDS1/geodata/geneseq/geneseqn/NA1993.DAT AAQ37481 +     6.00   114.61    205.15    2577  !
/SIDS1/geodata/geneseq/geneseqn/NA2000.DAT AAT17206 +     6.00   113.51    268.48    3602  !
/SIDS1/geodata/geneseq/geneseqn/NA1999.DAT AAX19202 +     6.00   113.39    246.03    3192  !
/SIDS1/geodata/geneseq/geneseqn/NA1998.DAT AAV11126 +     6.00   112.16    293.74    3757  !
/SIDS1/geodata/geneseq/geneseqn/NA1997.DAT AAT57537 +     6.00   123.66    816.92   12112  !
/SIDS1/geodata/geneseq/geneseqn/NA1997.DAT AAT57558 +     5.00   122.67     39.66      20  ! V
/SIDS1/geodata/geneseq/geneseqn/NA2000.DAT AAA75859 +     5.00   122.67     39.66      56  ! V
/SIDS1/geodata/geneseq/geneseqn/NA2000.DAT AAA75797 +     5.00   121.99     86.27      50  ! I
/SIDS1/geodata/geneseq/geneseqn/NA2001.DAT AAP54456 +     5.00   121.99     86.27      50  ! R
/SIDS1/geodata/geneseq/geneseqn/NA1997.DAT AAT71227 +     5.00   119.08    115.65      70  ! R
/SIDS1/geodata/geneseq/geneseqn/NA1997.DAT AAT49252 +     5.00   116.92    153.06      96  ! S
/SIDS1/geodata/geneseq/geneseqn/NA1997.DAT AAV75641 +     5.00   115.92    173.89     111  !
/SIDS1/geodata/geneseq/geneseqn/NA2000.DAT AAC16466 +     5.00   115.43    187.50     121  !
/SIDS1/geodata/geneseq/geneseqn/NA2000.DAT AAC16409 +     5.00   113.62    233.21     155  !
/SIDS1/geodata/geneseq/geneseqn/NA2000.DAT AAC22518 +     5.00   113.55    264.67     179  !
/SIDS1/geodata/geneseq/geneseqn/NA1997.DAT AAV34839 +     5.00   112.61    265.97     180  !
/SIDS1/geodata/geneseq/geneseqn/NA2000.DAT AAC23903 +     5.00   112.50    266.86     183  !
/SIDS1/geodata/geneseq/geneseqn/NA2000.DAT AAC27930 +     5.00   112.39    272.74     186  !
/SIDS1/geodata/geneseq/geneseqn/NA2000.DAT AAC20095 +     5.00   112.39    286.64     196  !
/SIDS1/geodata/geneseq/geneseqn/NA1999.DAT AAX68137 +     5.00   110.90    231.18     231  !
/SIDS1/geodata/geneseq/geneseqn/NA2000.DAT AAA45532 +     5.00   110.90    361.29     331  !
/SIDS1/geodata/geneseq/geneseqn/NA2000.DAT AAAR7142 +     5.00   110.44    342.49     340  !
/SIDS1/geodata/geneseq/geneseqn/NA1999.DAT AAX53801 +     5.00   110.33    356.25     351  !
/SIDS1/geodata/geneseq/geneseqn/NA1997.DAT AAT88164 +     5.00   110.22    361.24     255  !

seq_name: /SIDS1/geodata/geneseq/geneseqn/NA2000.DAT:AAA06617

seq_documentation_block:
ID   AAA06617 standard; cDNA; 278 BP.
XX
AC   AAA06617;
XX
DT   13-JUN-2000 (first entry)
XX
DE   Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:398.
XX
KW   Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW   immunogenic; cytostatic; vaccine; ss.
XX
OS   Homo sapiens.
XX
PN   WO200004149-A2.
XX
PD   27-JAN-2000.
XX
PF   14-JUL-1999;  99WO-US15838.
XX
PR   14-JUL-1998;  98US-0115453.
PR   14-JUL-1998;  98US-0116134.
PR   23-SEP-1998;  98US-0159812.
PR   23-SEP-1998;  98US-0159822.
PR   15-JAN-1999;  99US-0232149.
PR   15-JAN-1999;  99US-0232880.
PR   [illegible]-1999;  99US-0288046.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Dillman RO; Barkocy-Gallagher GA; Yuqin L; Xu J; Mitcham JL;
XX
DR   WPI; 2000-171289/15.
XX
PT   New polypeptide useful for treating and diagnosing prostate cancer
PT   comprises an immunogenic portion of prostate tumour protein -
XX
PS   Claim 50; Page 239; 263pp; English.
XX
CC   The present invention describes isolated polypeptides, comprising an
CC   immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC   and polynucleotides encoding them have cytostatic activity and can be
CC   useful in vaccines and in gene therapy. The polypeptides and
CC   polynucleotides encoding them, antigen presenting cells which express
CC   the polypeptides, antibodies against the polypeptides in vaccines
CC   comprising them can be used for inhibiting the development of prostate
CC   cancer in a patient. The polypeptides can be used to generate antibodies
CC   and/or idiotypic antibodies for passive immuno therapy. A portion of
CC   the polynucleotides encoding the polypeptides can be used as a probe or
CC   to modulate the expression of the polypeptides. AAA06241, AAA06691 and
CC   AAY82028 to AAY82029 represent sequences used in the exemplification of
CC   the present invention.
XX
SQ   Sequence 278 BP; 56 A; 86 C; 87 G; 49 T; 0 other;


alignment_scores:
                Quality:    9.00              Length:      9
                  Ratio:   1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_234_242 x AAA06617 ..
```

```
 Align seq 1/1  to: AAA06617  from: 1  to: 278

        1 SerTrpThrAspSerGluValAspSer 9
          |||||||||||||||||||||||||||
      149 AGCTGGACCGACAGCGAGGTGGACTCA 175

seq name: /cgn2/ftp/pir1/geneseq/geneseq-NA.a* AAI7AE:1/5

seq documentation block:
ID   AAE21828 standard; DNA; 1087 BP.
XX
AC   AAE21828;
XX
DT   27-MAR-2001  (first entry)
XX
DE   Human breast and ovarian cancer associated antigen gene SEQ ID 215.
XX
KW   Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
KW   nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW   antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
KW   antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
KW   Addison's disease; allergy; autoimmune haemolytic anaemia;
KW   autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW   multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW   cardiovascular disorder; wound healing; neurological disease; ds.
XX
OS   Homo sapiens.
XX
PN   WO200055178-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05881.
XX
PR   12-MAR-1999;  99US-0124270.
XX
PA   (HUMA ) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-615178/58.
DR   P-PSDB; AAB58925.
XX
PT   New human breast and ovarian cancer associated gene sequences and the
PT   polypeptides encoded by these genes, useful in the prevention,
PT   treatment and diagnosis of cancer, immune disorders, cardiovascular
PT   disorders and neurological diseases -
XX
PS   Claim 1; Page 642-643; 1299pp; English.
XX
CC   Sequences AAE21614 - AAE22031 represent DNA sequences encoding human
CC   proteins AAB58711 - AAB59128. The DNA and protein sequences are
CC   associated with breast and ovarian cancer. Included in the invention are
CC   sequences AAE22032 - AAE22040 and AAB59129 which are used in the
CC   isolation and characterisation of the DNA and protein sequences of the
CC   invention. The breast and ovarian cancer associated DNA, protein, agonist
CC   or antagonist sequences exhibit cytostatic, immunosuppressive,
CC   nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
CC   antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
CC   antibacterial; antifungal; antiparasitic and cardiant activity. The
CC   polynucleotide and protein sequences are used in the diagnosis of cancer,
CC   particularly breast and ovarian cancer. The nucleic acid sequences,
CC   proteins, agonists and agonists may also be used in the diagnosis,
CC   prevention and treatment of immune disorders e.g. Addison's disease,
CC   allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC   diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC   arthritis and ulcerative colitis, cardiovascular disorders such as
CC   myocardial ischaemias; wound healing; neurological diseases such as
CC   cerebral anoxia and epilepsy; and infectious diseases.
XX
SQ   Sequence 1087 BP; 215 A; 347 C; 339 G; 182 T; 4 other;

alignment scores:
            Quality:    9.000              Length:       9
              Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
     1 SerTrpThrAspSerGluValAspSer 9
       |||||||||||||||||||||||||||
   576 AGCTGGACCGACAGCGAGGTGGACTCA 602

seq name: /cgn2/ftp/pir1/geneseq/geneseq-NA2000.IA1:AAZ50691

seq documentation block:
ID   AAZ50691 standard; cDNA; 1894 BP.
XX
AC   AAZ50691;
XX
DT   23-MAY-2000  (first entry)
XX
DE   Human prostate derived Ets factor cDNA.
XX
KW   Prostate derived Ets factor; PDEF; human; chromosome 6p21.3; cancer;
KW   loss of heterozygosity; chromosomal translocation; linkage analysis;
KW   cytostatic; cardiant; immunosuppressive; cerebroprotective; fungicide;
KW   antibacterial; vulnerary; neuroprotective; antiparkinsonian; nootropic;
KW   anabolic; antiinflammatory; anorectic; hybridisation probe; forensic;
KW   tumour marker; diagnosis; treatment; prostate cancer; blood coagulation;
KW   autoimmune disorder; haematopoietic; immune/nervous system; stroke;
KW   neoplasm, microbial infection, tissue regeneration, heart attack;
KW   scarring; food additive; preservative; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             410..1423
FT                   /*tag= a
FT                   /product= "Human prostate-derived Ets factor"
FT                   /note= "Expressed in prostate epithelium"
XX
PN   WO200006589-A1.
XX
PD   10-FEB-2000.
XX
PF   02-AUG-1999;  99WO-US17470.
XX
PR   31-JUL-1998;  98US-0126945.
XX
PA   (HUMA ) HUMAN GENOME SCI INC.
PA   (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI   Libermann TA,  Oettgen HF,  Kunsch CA,  Endress GA,  Rosen CA;
XX
DR   WPI; 2000-156277-17.
DR   P-PSDB; AAY45004.
XX
PT   Novel prostate derived polypeptide, polynucleotide useful for
PT   diagnosis, prevention and treatment of prostate cancer, autoimmune
PT   disorders, microbial infections and also as food additive or
PT   preservative
XX
PS   Claim 4; Fig 1; 132pp; English.
XX
CC   The present cDNA sequence encodes a human prostate derived Ets factor
CC   (PDEF), isolated from human prostate carcinoma cDNA library
CC   (ATCC No.203057). PDEF gene expression is highest in tissues involved in
CC   androgen sensitivity, with enriched expression in prostate epithelium
CC   cells. PDEF gene is mapped to the human chromosome 6p21.3
CC   region that is associated with loss of heterozygosity and chromosomal
CC   translocations in various human cancers. PDEF has cytostatic, cardiant,
```

```
CC   immunosuppressive, cerebroprotective, fungicide, antibacterial,
CC   vulnerary, neuroprotective, antiparkinsonian, nootropic, anabolic,
CC   antiinflammatory and anorectic activity. PDEF polynucleotides are useful
CC   in linkage analysis as markers, as hybridisation probes for differential
CC   identification of the tissues or cell types and as polymorphic markers
CC   for forensic purposes. PDEF is useful as prostate-specific tumour marker
CC   for the diagnosis and treatment of prostate cancer. PDEF sequences are
CC   useful for treating autoimmune disorders, haematopoietic, blood
CC   coagulation, immune and nervous system disorders, hyperproliferative
CC   disorders like, neoplasms and microbial infections, heart attacks,
CC   stroke, scarring and for tissue regeneration. They are also useful as
CC   food additives or preservatives.
XX
SQ   Sequence 1894 BP; 468 A; 654 C; 571 G; 302 T; 0 other;


alignment_scores:
                Quality:    9.00              Length:       9
                  Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_234_242 x AAZ50691     .

 Align seq 1/1  to: AAZ50691  from: 1  to: 1894

        1 SerTrpThrAspSerGluValAspSer 9
          |||||||||||||||||||||||||
     1115 AGCTGGACCGACAGCGAGGTGGACTCA 1141

Seq_name: US-09-126-945B-2_COPY_234_242 x AAZ50691     .

seq_documentation_block:
ID   AAC84261 standard; DNA; 1894 BP.
XX
AC   AAC84261;
XX
DT   16-MAR-2001  (first entry)
XX
DE   Gland-specific Ets transcription factor (GSEF) nucleotide sequence.
XX
KW   Transcription factor; gland-specific Ets transcription factor; GSEF;
KW   metastatic potential; cancer; tumour; metastasis; breast; prostate;
KW   leukaemia; lymphoma; sarcoma; melanoma; chromosome 6p21.1-6p21.3; ds.
XX
OS   Homo sapiens.
XX
PN   WO200070092-A1.
XX
PD   23-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US13173.
XX
PR   14-MAY-1999;   99US-0134112.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Kaufmann J,  Xin H,  Harrowe G;
XX
DR   WPI; 2001-041019/05.
DR   P-PSDB; AAB49628.
XX
PT   Detecting metastatic and potential metastatic cancerous cells, useful
PT   for diagnosing, prognosing, grading and staging of cancers by detecting
PT   gland-specific Ets transcription factor gene product in a biological
PT   sample from a cell -
XX
PS   Disclosure; Page 81-83; 95pp; English.
XX
CC   This invention relates to a method for the detection and determination of
CC   the metastatic potential of a cell. The method comprises detecting a
CC   gland-specific Ets transcription factor (GSEF) gene product in a test
CC   sample. Detection of a GSEF gene product in the test sample in amount
CC   lower than that in a normal cell, is indicative of a cell with high
CC   metastatic potential. The method is useful for determining the metastatic
CC   potential of a cell, for the diagnosis and prognosis of cancer as well as
CC   grading and staging of cancers by detecting GSEF expression in a
CC   biological test sample. The method may also be used to monitor patients
CC   having a predisposition to develop a particular cancer. GSEF polypeptides
CC   are useful for producing antibodies, in cancer diagnosis, prognosis,
CC   grading, staging and management of breast and prostate tumours, and in
CC   detecting polymorphisms in the sequence. GSEF genes and proteins are also
CC   useful in gene therapy. GSEF gene product expression levels can be used
CC   in conjunction with any tissue in which an alteration in GSEF gene
CC   product expression levels is associated with development of a
CC   cancer-associated phenotype. Cancers, which can be monitored include
CC   cancers of the prostate, cervix, lung and colon, melanomas, colorectal
CC   adenocarcinomas, Wilm's tumour, retinoblastoma, sarcomas, myosarcomas,
CC   lung carcinomas, leukaemia, and lymphomas. The GSEF gene is located on
CC   human chromosome 6, specifically at 6p21.1-6p21.3. The present sequence
CC   represents the DNA encoding GSEF.
XX
SQ   Sequence 1894 BP; 468 A; 654 C; 571 G; 302 T; 0 other;


alignment_scores:
                Quality:    9.00              Length:       9
                  Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity:  100.000

alignment_block:
 US-09-126-945B-2_COPY_234_242 x AAC84261     .

 Align seq 1/1  to: AAC84261  from: 1  to: 1894

        1 SerTrpThrAspSerGluValAspSer 9
          |||| ||||||||||||||||||||||
     1115 AGCTGGACCGACAGCGAGGTGGACTCA 1141

Seq_name: US-09-126-945B-2_COPY_234_242 x AAC84261     .

seq_documentation_block:
ID   AAC84266 standard; DNA; 3317 BP.
XX
AC   AAC84266;
XX
DT   16-MAR-2001  (first entry)
XX
DE   Gland-specific Ets transcription factor (GSEF) cDNA sequence.
XX
KW   Transcription factor; gland-specific Ets transcription factor; GSEF;
KW   metastatic potential; cancer; tumour; metastasis; breast; prostate;
KW   leukaemia; lymphoma; sarcoma; melanoma; chromosome 6p21.1-6p21.3; ss.
XX
OS   Homo sapiens.
XX
PN   WO200070092-A1.
XX
PD   23-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US13173.
XX
PR   14-MAY-1999;   99US-0134112.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Kaufmann J,  Xin H,  Harrowe G;
XX
DR   WPI; 2001-041019/05.
DR   P-PSDB; AAB49633.
XX
PT   Detecting metastatic and potential metastatic cancerous cells, useful
PT   for diagnosing, prognosing, grading and staging of cancers by detecting
PT   gland-specific Ets transcription factor gene product in a biological
PT   sample from a cell
XX
```

PS Disclosure; Fig 2; 95pp; English.
XX
CC This invention relates to a method for the detection and determination of
CC the metastatic potential of a cell. The method comprises detecting a
CC gland specific Ets transcription factor (GSEF) gene product in a test
CC sample. Detection of a GSEF gene product in the test sample in amount
CC lower than that in a normal cell, is indicative of a cell with high
CC metastatic potential. The method is useful for determining the metastatic
CC potential of a cell, for the diagnosis and prognosis of cancer as well as
CC grading and staging of cancers by detecting GSEF expression in a
CC biological test sample. The method may also be used to monitor patients
CC having a predisposition to develop a particular cancer. GSEF polypeptides
CC are useful for producing antibodies, in cancer diagnosis, prognosis,
CC grading, staging and management of breast and prostate tumours, and in
CC detecting polymorphisms in the sequence. GSEF genes and proteins are also
CC useful in gene therapy. GSEF gene product expression levels can be used
CC in conjunction with any tissue in which an alteration in GSEF gene
CC product expression levels is associated with development of a
CC cancer associated phenotype. Cancers, which can be monitored include
CC cancers of the prostate, cervix, lung and colon, melanomas, colorectal
CC adenocarcinomas, Wilm's tumour, retinoblastoma, sarcomas, myosarcomas,
CC lung carcinomas, leukaemia, and lymphomas. The GSEF gene is located on
CC human chromosome 6, specifically at 6p21.1 6p21.3. The present sequence
CC represents the cDNA encoding GSEF
XX
SQ Sequence 3417 BP; 710 A; 1026 C; 970 G; 611 T; 0 other;

alignment scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2_COPY_234_242 x AAC84266 .

Align seq 1/1 to: AAC84266 from: 1 to: 3417

1 SerTrpThrAspSerGluValAspSer 9
||||||||||||||||||||||||||||
2514 AGCTGGACCCACAGCGAGGTGGACTCA 2540

Seq name: /cgn2_6/ptodata/1/pubseq/pubseq_os/pg_NA1995.DAT:AAC84266

seq documentation block:
ID AAC84266 standard; cDNA to mRNA; 3417 BP.
XX
AC AAC84266;
XX
DT 25-SEP-1996 (first entry)
XX
DE Human gene signature HUMGS005520.
XX
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
XX
OS Homo sapiens.
XX
PN WO9514772-A1.
XX
PD 01-JUN-1995.
XX
PF 11-NOV-1994; 94WO-JP01916.
XX
PR 12-NOV-1993; 93JP-0355504.
XX
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
XX
PI Matsubara K, Okubo K;
XX
DR WPI; 1995-206941/27.
XX
PT Identifying set gene signatures in 3' directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
XX
PS Claim 1; Page 1620; 2245pp; Japanese.
XX
CC A single stranded DNA (or its complementary strand or the corresp.
CC double stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in AAI19001-126837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
XX
SQ Sequence 395 BP; 99 A; 75 C; 86 G; 115 T; 20 other;

alignment scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-126-945B-2_COPY_234_242 x AAT24481/rev .

Align seq 1/1 to reverse of: AAT24481 from: 1 to: 395

2 TrpThrAspSerGluVal 7
|||||||||||||||||
389 TGGACAGACTCAGAGGTA 422

Seq name: /cgn2_6/ptodata/1/pubseq/pubseq_os/pg_NA2000.DAT:AAC49470

seq documentation block:
ID AAC49470 standard; DNA; 426 BP.
XX
AC AAC49470;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 61280.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132864.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135354.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136362.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148841.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151503.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158869.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.

```
PR   14 OCT 1999;   99US-0159329.
PR   14 OCT 1999;   99US-0159330.
PR   14 OCT 1999;   99US-0159331.
PR   14 OCT 1999;   99US-0159637.
PR   14 OCT 1999;   99US-0159638.
PR   18 OCT 1999;   99US-0159584.
PR   21 OCT 1999;   99US 0160741.
PR   21 OCT 1999;   99US-0160767.
PR   21 OCT 1999;   99US-0160768.
PR   21 OCT 1999;   99US-0160770.
PR   21 OCT 1999;   99US-0160814
PR   21 OCT 1999;   99US-0160815.
PR   22 OCT 1999;   99US-0160980.
PR   22 OCT 1999;   99US-0160981.
PR   22 OCT 1999;   99US-0160989.
PR   25 OCT 1999;   99US-0161404.
PR   25 OCT 1999;   99US-0161405.
PR   25 OCT 1999;   99US-0161406.
PR   26 OCT 1999;   99US-0161359.
PR   26 OCT 1999;   99US-0161360.
PR   26 OCT 1999;   99US-0161361.
PR   28 OCT 1999;   99US 0161920.
PR   28 OCT 1999;   99US 0161992.
PR   28 OCT 1999;   99US 0161993.
PR   29 OCT 1999;   99US-0162142.

Alignment scores:
              Quality:    6.00              Length:      6
                Ratio:   1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-126-945B-2-COPY_234_242 x AAC49470   ..

 Align seq 1/1  to: AAC49470  from: 1  to: 426

        3 ThrAspSerGluValAsp 8
          ||||||||||||||||||
      322 ACAGATTCTGAAGTTGAT 339

Seq name: /cds1/gcgdata/genseqp/genseqp_NA_PUB.DAT:AAC59536

Seq documentation block:
ID   AAC59536 standard; cDNA; 687 BP.
XX
AC   AAC59536;
XX
DT   15-FEB-2001  (first entry)
XX
DE   Human secreted protein cDNA sequence #30.
XX
KW   Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW   antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW   vulnerary, anticonvulsant, antibacterial, antifungal, antiparasitic,
KW   cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; human; secreted protein; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055352-A2.
XX
PD   21-SEP-2000.
XX
PF   09-MAR-2000; 2000WO-US06044.
XX
PR   12-MAR-1999;   99US-0124099.
PR   03-DEC-1999;   99US-0168661.
XX
PA   (HUMA ) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM,  Komatsoulis G;
XX
DR   WPI; 2000-602124/57.
DR   P-PSDB; AAB34245.
XX
PT   Novel human secreted proteins useful for diagnosis, prevention and
PT   treatment of disorders including neurological, cell proliferative,
PT   cardiovascular, autoimmune and inflammatory disorders and microbial
PT   infections
XX
PS   Claim 1; Page 345; 483pp; English.
XX
CC   The invention relates to the isolation of genes AAC59507-C59556 encoding
CC   50 human secreted proteins AAB34219-B34264). The genes can be used to
CC   generate fusion proteins by linking to the gene for the human
CC   immunoglobulin G Fc portion (AAC59498) for increasing the stability of
CC   the fusion protein as compared to the human protein only. The genes and
CC   proteins are useful for preventing, ameliorating or treating medical
CC   conditions, e.g. by protein or gene therapy. The genes are isolated
CC   from a range of human tissues disclosed in the specification. The
CC   nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC   the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer, and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis, (c) cardiovascular disorders such as myocardial ischaemias; (d)
CC   wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
XX
SQ   Sequence 687 BP; 174 A; 108 C; 114 G; 290 T; 1 other;


Alignment scores:
              Quality:    6.00              Length:      6
                Ratio:   1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-126-945B-2_COPY_234_242 x AAC59536/rev   ..

 Align seq 1/1  to: reverse of: AAC59536  from: 1  to: 687

        2 TrpThrAspSerGluVal 7
          |||| ||||||||||||
      262 TGGACTGATAGTGAAGTT 245

Seq name: /cds1/gcgdata/genseqp/genseqp_NA_PUB.DAT:AAF07726

Seq documentation block:
ID   AAF07726 standard; cDNA; 716 BP.
XX
AC   AAF07726;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Fusarium venenatum EST SEQ ID NO:249.
XX
KW   Multiple gene expression; filamentous fungal cell; EST;
KW   expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW   Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW   culture condition; environmental stress; spore morphogenesis;
KW   metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS   Fusarium venenatum.
XX
PN   WO200056762-A2.
XX
PD   28-SEP-2000.
XX
PF   22-MAR-2000; 2000WO-US07781.
XX
PR   22-MAR-1999;   99US-0273623.
XX
```

PA (NOVO) NOVO NORDISK BIOTECH INC.
PA (NOVO) NOVO NORDISK AS.
XX
PT Borka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
DR WPI; 2000-594572/56.
XX
PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags
XX
PS Claim 86; Page 478; 3161pp; English.
XX
CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.
XX
SQ Sequence 716 BP; 172 A; 214 C; 172 G; 157 T; 1 other;

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_234_242 x AAF07726 .

Align seq 1/1 to: AAF07726 from: 1 to: 716

3 ThrAspSerGluValAsp 8
612 ACCGACAGCGAAGTCGAC 629

seq_name: /SIBel/prodata/geneseq/geneseq_NA/GENESEQ.DAT:AAF64219

seq_documentation_block:
ID AAF64219 standard; cDNA; 1007 BP.
XX
AC AAF64219;
XX
DT 06-APR-2001 (first entry)
XX
DE Human secreted protein gene 44 SEQ ID NO:54.
XX
KW Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KW nootropic; neuroprotective; antibacterial; virucide; fungicide;
KW ophthalmological; vulnerary; autoimmune disease; cardiovascular disorder;
KW hyperproliferative disorder; cerebrovascular disorder; wound healing;
KW nervous system disorder; ocular disorder; skin ageing; chemotaxis;
KW food additive; ss.
XX
OS Homo sapiens.
XX
PN WO200077026-A1.
XX
PD 21 DEC 2000.
XX
PF 01 JUN 2000; 2000WO-US14973.
XX
PR 11 JUN 1999; 99US-0138630.
XX
PA (HUMA) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Rosen CA, Ruben SM, Komatsoulis GA;
XX
DR WPI; 2001-071258/08.
DR P-PSDB; AAB75549.
XX
PT Nucleic acid molecules encoding human secreted proteins, used in
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
PT Parkinson's diseases and cancers -
XX
PS Claim 1; Page 465; 542pp; English.
XX
CC Human secreted proteins AAB75506 - AAB75554 are encoded by polynucleotide
CC sequences AAF64176 - AAF64224. The specification includes amino acid
CC sequences AAB75555 - AAB75606 which represent fragments of the human
CC secreted proteins, and protein sequences with which they share homology.
CC The proteins and polynucleotides, their agonists and antagonists have
CC activities dependent on the tissues and cells in which they are
CC expressed, examples of these activities include, immunosuppressive;
CC antiarthritic, antirheumatic, antiproliferative, cytostatic, cardiant;
CC vasotropic, cerebroprotective, nootropic, neuroprotective; antibacterial;
CC virucide; fungicide; opthalmalogical; and vulnerary. The proteins,
CC polynucleotides, agonists and antagonists can be used to treat or detect
CC or diagnose various diseases and disorders including, autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders
CC e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,
CC infections caused by bacteria, viruses and fungi and ocular disorders
CC e.g. corneal infection. The polypeptides can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin ageing due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities. Included in the invention are
CC polynucleotide sequences AAF64167 - AAF64175 and peptide AAB75505 which
CC are used in the isolation, identification and characterisation of the
CC proteins of the invention.
XX
SQ Sequence 1007 BP; 251 A; 181 C; 173 G; 400 T; 2 other;

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_234_242 x AAF64219/rev ..

Align seq 1/1 to reverse of: AAF64219 from: 1 to: 1007

2 TrpThrAspSerGluVal 7
582 TGGACTGATAGTGAAGTT 565

seq_name: /SIBel/prodata/geneseq/geneseq_NA/GENESEQ.DAT:AAN80229

seq_documentation_block:
ID AAN80229 standard; DNA; 1236 BP.

```
XX
AC   AAN80229;
XX
DT   03-DEC-1990  (first entry)
XX
DE   Bioadhesive precursor protein gene cDNA 52.
XX
KW   Bioadhesive precursor protein; cDNA 52; mussels; barnacles; oysters;
KW   biomedical adhesive; sealants; wound healing; ss.
XX
OS   Mytilus edulis
XX
FH   Key             Location/Qualifiers
FT   CDS             3..1004
FT                   /*tag= a
FT                   /product bioadhesive precursor protein
XX
PN   WO8803954-A.
XX
PD   02-JUN-1988.
XX
PF   24-NOV-1987;   87WO-US03048.
XX
PR   07-AUG-1987;   87US-0082456.
XX
PA   (GENE ) GENEX CORP.
XX
PI   Waugh KJ,  Anderson DM,  Strausberg R,  Strausberg SL,  McCandliss R;
XX
DR   WPI; 1988-161622/23.
DR   P-PSDB; AAP82971.
XX
PT   DNA coding for bio-adhesive precursor protein - obtd. from mussels,
PT   barnacles and oysters and used exp. to prepare biomedical adhesive.
XX
PS   Disclosure; 7 pp; English.
XX
CC   mRNA from the phenol gland of the foot of M. edulis is isolated for
CC   the synthesis of cDNA.  The cDNA is inserted into a bacteriophage
CC   and E. coli transfected and cultured.  Five clones were isolated and
CC   characterised, one of which is clone 52.  The sequence has been
CC   divided to show 26 subunits, including 23 decapeptides and 3
CC   hexapeptides, and an amino terminal proline rich segment.
CC   There is also a direct repeat of DNA sequences in clone 52.
CC   This cDNA clone and two of the other (55 (AAN80230) and 56 (AAN80231))
CC   contain some noncoding DNA sequence (not shown) which is believed
CC   to be derived from intron sequences.
CC   Clone 14-1 (AAN80228) and 52 display extensive homology at their 3'
CC   ends. In particular, the last 138 codons of the translated regions are
CC   identical and include codons for a hexapeptide followed by 12
CC   decapeptides (beginning at codon 75 of clone 14-1 and codon 205
CC   of clone 52.  The DNA segment contg. the coding regions of
CC   bioadhesive precursor protein of clone 14-1 and 52 was obtained from
CC   lambda gt10 cloning vectors by EcoRI digestion and subcloned into the
CC   EcoRI site of plasmid pUC26-27.
CC   The protein is chemically or enzymatically hydroxylated to produce
CC   adhesive composn. having excellent properties in wet environments. The
CC   product is partic. useful as a biomedical adhesive or sealant e.g. in
CC   wound healing or as a marine or dental adhesive.  It can be used as a
CC   thin film membrane.
CC   See also AAN80228-44.
XX
SQ   Sequence 1236 BP; 464 A; 338 C; 125 G; 309 T; 0 other;


alignment scores:
             Quality:    6.00              Length:       6
               Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment Block:
@US-09-126-945B-2-COPY_234_242-x-AAN80229/rev
```

```
 Align seq 1/1  to reverse of: AAN80229  from: 1  to: 1236

       4 AspSerGlnValAspSer 9
         ||-| ||||||||||||
     204 GAC TCC AAG TCG GAC TCG 187

sequence: /*  = 1/1: file: /usr/seq/genseq/NA1998.DAT:AAN82450

Seq documentation block:
ID   AAN82450 standard; cDNA; 1236 BP.
XX
AC   AAN82450;
XX
DT   06-MAR-1992  (first entry)
XX
DE   Sequence for a gene identified as cDNA clone 52 which codes for
DE   a bioadhesive precursor protein.
XX
KW   Bioadhesive; wound healing; bonding; recombinant adhesive; ss.
XX
OS   Mytilus edulis.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..1022
FT                   /*tag= a
XX
PN   WO8807076-A.
XX
PD   22-SEP-1988.
XX
PF   11-MAR-1988;   88WO-US00876.
XX
PR   12-MAR-1987;   87US-0025243.
XX
PA   (GENE ) GENEX CORP.
XX
PI   Waugh KJ,  Anderson DM,  Strausberg SL,  Strausberg R;
PI   Wei T;
XX
DR   WPI; 1988-285533/40.
DR   P-PSDB; AAP83194.
XX
PT   Bio-adhesive precursor protein analogues prodn.  by DNA
PT   recombinant techniques, subsequent hydroxylation, gives an
PT   adhesive for use in wet environment
XX
PS   Example; Fig 14; 101pp; English.
XX
CC   The bioadhesive precursor protein analogue of the invention can be
CC   hydroxylated (prod. claimed) and used as an adhesive in wet
CC   environments. The resulting bioadhesives have marine, biomedical and
CC   dental applications. They may be used e.g. in wound healing in the
CC   same manner as fibrin. While the decapeptide ala-lys-pro-ser-tyr-pro-
CC   pro-thr-tyr-lys is repeated many times in the polyphenolic adhesive
CC   protein of M.edulis, the examination of cDNA sequences encoding
CC   portions of this protein (see U.S. Patent Application Serial No. 943,
CC   945) reveals that many other repetitive decapeptide and hexapeptide
CC   sequences are also present in the protein, and that these other
CC   sequences may constitute the majority of AA sequence of the
CC   polyphenolic adhesive protein of M.edulis. For example, in cDNA
CC   clone 14-1 nineteen decapeptides and one hexapeptide are encoded.
XX
SQ   Sequence 1236 BP; 462 A; 340 C; 124 G; 310 T; 0 other;


alignment scores:
             Quality:    6.00              Length:       6
               Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment Block:
 US-09-126-945B-2-COPY_234_242-x-AAN82450/rev
```

Align seq 1/1 to reverse of: AAN82450 from: 1 to: 1236

4 AspSerGluValAspSer 9
|||||||||||||||||||
204 GACTCGGAGGTGGACTCG 187

Seq_name: "????" in file:/data/geneseq-prot/pDNA2_1.fsa:AAF14511

Seq_documentation_block:
ID AAF14511 standard; cDNA; 1617 BP.
XX
AC AAF14511;
XX
DT 13-MAR-2001 (first entry)
XX
DE Aspergillus oryzae EST SEQ ID NO:7044
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Aspergillus oryzae.
XX
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US007781.
XX
PR 22-MAR-1999; 99US-0273623.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
PA (NOVO) NOVO NORDISK AS.
XX
PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
DR WPI; 2000-594572/56.
XX
PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags
XX
PS Claim 88; Page 2856; 3161pp; English.
XX
CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC AAF14879 to AAF15437 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.
XX
SQ Sequence 1617 BP; 446 A; 378 C; 381 G; 412 T; 0 other;

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-126-945B-2 (COPY_234_242 x AAF14511 ..

Align seq 1/1 to: AAF14511 from: 1 to: 1617

3 ThrAspSerGluValAsp 8
||||||||||||||||||
462 ACCGACAGTGAAGTTGAT 479

Seq_name: "????" in file:/data/geneseq-prot/pDNA2_1.fsa:AAX88977

Seq_documentation_block:
ID AAX88977 standard; DNA; 2044 BP.
XX
AC AAX88977;
XX
DT 16-SEP-1999 (first entry)
XX
DE Mouse dephosphorylase inhibiting p91-like protein #3 encoding DNA.
XX
KW Dephosphorylase inhibiting protein; p91; tyrosine phosphatase SHP 1;
KW SHP-2; inositol-5-phosphate SHIP; phosphating tyrosine;
KW immunoreceptor; immunomodulatory agent; ss.
XX
OS Mus sp
XX
PN JP11169184 A.
XX
PD 29-JUN-1999.
XX
PF 12-DEC-1997; 97JP-0362285.
XX
PR 12-DEC-1997; 97JP-0362285.
XX
PA (UYOK) UNIV OKAYAMA.
XX
DR WPI; 1999-422622/36.
DR P-PSDB; AAY24320.
XX
PT New peptide - useful for inhibiting dephosphorylase
XX
PS Claim 7; Page 19-20; 30pp; Japanese.
XX
CC The present invention describes new proteins for inhibiting
CC dephosphorylase. The proteins can be combined with tyrosine phosphatase
CC SHP-1, SHP-2 or inositol-5 phosphatase SHIP by phosphating tyrosine.
CC The present invention also describes an immunoreceptor comprising one
CC the above proteins, and DNA coding the above proteins. The new proteins
CC can be used as an immunomodulatory agent. The present sequence encodes
CC a protein from the present invention.
XX
SQ Sequence 2044 BP; 556 A; 577 C; 457 G; 454 T; 0 other;

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_234_242 x AAX88977 ..

Align seq 1/1 to: AAX88977 from: 1 to: 2044

1 SerIleThrAspSerGlu 6
||||||||||||||||||
30 TCTTGGACTGACTCTGAG 47

```
Seq name: /cygdrive/c/geneata/geneseq-pro/seqs/NA/US07/US07_AAN70807

Seq documentation block:
ID   AAN70807 standard; DNA; 2440 BP.
XX
AC   AAN70807;
XX
DT   02-APR-1991  (first entry)
XX
DE   Sequence encoding the NA4 antigen of Eimeria necatrix (genomic
DE   clone).
XX
KW   Poultry vaccine; Eimeria tenella vaccine; Eimeria necatrix vaccine;
KW   ss.
XX
OS   Eimeria necatrix.
XX
FH   Key             Location/Qualifiers
FT   exon            664..754
FT                   /*tag   a
FT   exon            856..1028
FT                   /*tag   b
FT   exon            1151..1512
FT                   /*tag   c
FT   exon            1630..1750
FT                   /*tag   d
XX
PN   AU8665869-A.
XX
PD   04-JUN-1987.
XX
PF   01-DEC-1986;    86AU-006586 9.
XX
PR   11-DEC-1985;    85US-0808013.
PR   03-DEC-1985;    85US-0805401.
PR   06-DEC-1985;    85US-0805824.
PR   11-DEC-1985;    85US-0807497.
XX
PA   (SOLV ) SOLVAY & CIE.
XX
PI   Newman KA,  Gore JC,  Tedesco JL,  Peterson GR,  Brothers VM;
PI   Paul LS,  Chang RJ,  Andrews WH;
PI   Kuhn I,  McCaman MT,  Files JG,  Silas SR;
XX
DR   WPI; 1987-199027/29.
DR   P-PSDB; AAP70490.
XX
PT   New polypeptide for protecting chickens against coccidiosis - are
PT   obtd. by recombinant DNA procedures involving total genomic DNA
PT   from eimeria maxima oocysts.
XX
PS   Disclosure; Fig 17; 255pp; English.
XX
CC   Eimeria tenella polypeptides are used to immunise chickens agains
CC   infection against Eimeria strains, so that coccidiosis is prevent.
CC   They are used in vaccines, e.g. in a dose to provide over 20
CC   microgram/kg. The monoclonal antibody or its anti-idiotypic antibody
CC   may also be administered to confer protection. The sequences in Fig
CC   4 (AAN70804, AAP70487) are very similar (identical apart from DNA
CC   ambiquities?) to the sequences in Fig 5 (AAN70805, AAP70488).
XX
SQ   Sequence 2440 BP; 598 A; 584 C; 591 G; 666 T; 2 other;


Alignment scores:
              Quality:     6.00              Length:       6
                Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

Alignment block:
#US-09-126-945B-2-COPY-234-242 x AAN70807/rev  ..

        .
```

```
Align seq 1/1  to reverse of: AAN70807  from: 1  to: 2440

       4 AspSerGluValAspSer 9
         ||||||||||||||||||
     513 GACAGCCAGCTCGACAGC 526
```

(M of: US-09-126-945B-2_COPY_234_242 [redacted]

Date: Sep 11, 2001 9:58 PM

About: Results were produced by the Genetics Analysis version 1.6
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL framer_p2s model -DEV xlp
-Q [redacted]
-DB Issued_Patents_NA -QFMT fastap -SUFFIX olip2n.rni
-GAPOP 4.500 -GAPEXT 0.050 -MINMATCH 0.100 -LOOPCL 0.000
-LOOPEXT 0.000 -QGAPOP 4.500 -QGAPEXT 0.050 -XGAPOP 60.000
-XGAPEXT 60.000 -PGAPOP 6.000 -PGAPEXT 7.000 -YGAPOP 60.000
-YGAPEXT 60.000 -DELOP 6.000 -DELEXT 7.000 -START 1
-MATRIX oligo -TRANS human40.cdi -LIST 45 -DOALIGN 200
-THR_SCORE quality -THR_MIN 1 -ALIGN 15 -MODE LOCAL -OUTFMT pir
-NORM ext -MINLEN 0 -MAXLEN 2000000000
-USER [redacted]
-WAIT -THREADS 1

Search information block:
Query: US-09-126-945B-2_COPY_234_242
Query length: 9
Database: Issued_Patents_NA:*
Database sequences: 324599
Database length: 94655562
Search time (sec): 322.720000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000

score_list:
Sequence Strd Orig ZScore EScore Len ! Documentation ..
[redacted]

seq_no = [redacted] -data /cgn/backfile1 seq=5202236-39

Seq_documentation_block:
;Patent No: 5202236
; APPLICANT: MARCH, KATHY J.;ANDERSON, DAVID M.;STRAUSBERG,
;SUSAN L.;MCCANDLISS, RUSS;WEI, TENA;FILPULA, DAVID
; TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE
;PROTEIN
; NUMBER OF SEQUENCES: 39
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/529,762
; FILING DATE: 25-MAY-1990
; APPLICATION NUMBER: 82,456
; FILING DATE: 07 AUG-1987
; APPLICATION NUMBER: 933,945
; FILING DATE: 24-NOV-1986
; APPLICATION NUMBER: 650,128
; FILING DATE: 13-SEP-1984
;SEQ ID NO:39;
; LENGTH: 994
5202236-39

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

[redacted]

Alignment of: 1 to [redacted] 5202236-39 from: 1 to: 994

1 AspGlnGlnValAspSer 9
[redacted]
994 GATCGTGGTGGACTGG 177

seq_no = [redacted] -data /cgn/backfile1 seq=5202236 4

Seq_documentation_block:
;Patent No: 5202236
; APPLICANT: MARCH, KATHY J.;ANDERSON, DAVID M.;STRAUSBERG,
;SUSAN L.;MCCANDLISS, RUSS;WEI, TENA;FILPULA, DAVID
; TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE
;PROTEIN
; NUMBER OF SEQUENCES: 39
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/529,762
; FILING DATE: 25-MAY-1990
; APPLICATION NUMBER: 82,456
; FILING DATE: 07 AUG-1987
; APPLICATION NUMBER: 933,945
; FILING DATE: 24-NOV-1986
; APPLICATION NUMBER: 650,128
; FILING DATE: 13-SEP-1984
;SEQ ID NO:4;
; LENGTH: 1276
5202236-4

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```
alignment block:
 US-09-126-945B-2_COPY_234_242 x 5202236.3/rev   ..

 Align seq 1/1  to reverse of: 5202236.4  from: 1  to: 1236

        4 AspSerGlnValAspSer 9
          ||||||||||||||||||
      204 GACTCGGAGGTGGACTCG 187

seq name: /cgi-data/Test/A COMB.seq:... ... ... 878 18

seq documentation block:
; Sequence 18, Application US/08490878
; Patent No. 5700683
;  GENERAL INFORMATION:
;     APPLICANT:  Stover, Charles K.
;     APPLICANT:  Mahairas, Gregory G.
;     TITLE OF INVENTION:  VIRULENCE ATTENUATING GENETIC DELETIONS
;     NUMBER OF SEQUENCES:  18
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Townsend and Townsend Khourie and Crow
;       STREET:  One Market Plaza, Steuart Street Tower, 20th
;       STREET:  Floor
;       CITY:  San Francisco
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  94105
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/390,878
;       FILING DATE:  17 FEB 1995
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Hunter, Tom
;       REGISTRATION NUMBER:  38,498
;       REFERENCE/DOCKET NUMBER:  15371A-17
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415/543/9600
;       TELEFAX:  415/543/5043
;   INFORMATION FOR SEQ ID NO:  18:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  12412 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-490-878-18


alignment scores:
             Quality:     6.00              Length:       6
               Ratio:  1.000                Gaps:         0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_234_242 x US-08-490-878-18/rev   ..

 Align seq 1/1  to reverse of: US-08-490-878-18  from: 1  to: 12412

        4 AspSerGlnValAspSer 9
          ||||||||||||||||||
     1459 GATTCGGAGGTAGACAGC 1442

seq name: /cgi-data/Test/A COMB.seq:US-08-979-269-24

seq documentation block:
; Sequence 24, Application US/08979269
; Patent No. 6037127
;  GENERAL INFORMATION:
;     APPLICANT:  EBERSOLE, RICHARD C.
;     APPLICANT:  BENCEN-KESSEN, EDWIN R.
;     APPLICANT:  MAJARIAN, WILLIAM R.
;     APPLICANT:  MCELLIGOTT, SANDRA G.
;     APPLICANT:  PAYNE, MARK S.
;     APPLICANT:  RAFALSKI, JAN A.
;     TITLE OF INVENTION:  A NOVEL METHOD FOR DETECTION
;     TITLE OF INVENTION:  OF NUCLEIC ACID FRAGMENTS
;     NUMBER OF SEQUENCES:  27
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  E. I. DU PONT DE NEMOURS AND COMPANY
;       STREET:  1007 MARKET STREET
;       CITY:  WILMINGTON
;       STATE:  DELAWARE
;       COUNTRY:  U.S.A.
;       ZIP:  19898
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  DISKETTE, 3.50 INCH
;       COMPUTER:  IBM PC COMPATIBLE
;       OPERATING SYSTEM:  MICROSOFT WINDOWS 95
;       SOFTWARE:  MICROSOFT WORD VERSION 7.0A
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/979,269
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/221,769
;       FILING DATE:  MARCH 31, 1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  SIEGELL, BARBARA C.
;       REGISTRATION NUMBER:  30,684
;       REFERENCE/DOCKET NUMBER:  CR-9422-C
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  302-992-4931
;       TELEFAX:  302-773-0164
;   INFORMATION FOR SEQ ID NO:  24:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  20 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  other nucleic acid
;     DESCRIPTION:  /desc= "VEE 5' primer"
US-08-979-269-24


alignment scores:
             Quality:     5.00              Length:       5
               Ratio:  1.000                Gaps:         0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-126-945B-2_COPY_234_242 x US-08-979-269-24/rev   ..

 Align seq 1/1  to reverse of: US-08-979-269-24  from: 1  to: 20

        5 SerGlnValAspSer 9
          |||||||||||
       19 TCAGAAGTCGACAGC 5

Seq name: /cgi-data/Test/A COMB.seq:US-08-484-322-27

seq documentation block:
; Sequence 27, Application US/08484322
; Patent No. 5710015
;  GENERAL INFORMATION:
;     APPLICANT:  BUKH, J., MILLER, R.H. AND
;     APPLICANT:  PURCELL, R.H.
;     TITLE OF INVENTION:  NUCLEOTIDE AND DEDUCED
;     TITLE OF INVENTION:  AMINO ACID SEQUENCES OF HYPERVARIABLE
;     TITLE OF INVENTION:  REGION 1 OF THE ENVELOPE 2 GENE OF ISOLATES
;     TITLE OF INVENTION:  OF HEPATITIS C VIRUS AND THE USE OF
;     TITLE OF INVENTION:  REAGENTS DERIVED FROM THESE HYPERVARIABLE
```

```
;     TITLE OF INVENTION:  SEQUENCES IN DIAGNOSTIC METHODS AND
;     TITLE OF INVENTION:  VACCINES
;     NUMBER OF SEQUENCES:  98
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  MORGAN & FINNEGAN
;       STREET:  345 PARK AVENUE
;       CITY:  NEW YORK
;       STATE:  NEW YORK
;       COUNTRY:  USA
;       ZIP:  10154
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  FLOPPY DISK
;       COMPUTER:  IBM PC COMPATIBLE
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  WORDPERFECT 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/484,822
;       FILING DATE:  June 7, 1995
;       CLASSIFICATION:  424
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  RICHARD W. BORK
;       REGISTRATION NUMBER:  36,459
;       REFERENCE/DOCKET NUMBER:  2026-411P95
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 758-4800
;       TELEFAX:  (212) 751-6849
;       TELEX:  421792
;   INFORMATION FOR SEQ ID NO:  22:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  96 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     ORIGINAL SOURCE:
;       ORGANISM:  homosapiens
;       INDIVIDUAL ISOLATE:  S9
US-08-484-822-22


alignment_scores:
            Quality:    5.00              Length:      5
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_234_242 x US-08-484-822-22     .

 Align seq 1/1  to: US-08-484-822-22  from: 1  to: 96

       2 TrpThrAspSerGlu 6
         ||||||||||||||||
      66 TGGACCGACTCAGAA 80

seq_name: /tmp2_6/pi.bin/2_iea/5A_COMB.seq ** # 136.586.1

seq_documentation_block:
; Sequence 4, Application US/08246583
; Patent No. 5750394
;  GENERAL INFORMATION:
;     APPLICANT:  Palese, Peter
;     APPLICANT:  O'Neill, Robert
;     TITLE OF INVENTION:  IDENTIFICATION AND USE OF ANTIVIRAL
;     TITLE OF INVENTION:  COMPOUNDS THAT INHIBIT INTERACTION OF HOST CELL PROTEINS
;     TITLE OF INVENTION:  AND VIRAL PROTEINS REQUIRED FOR VIRAL REPLICATION
;     NUMBER OF SEQUENCES:  15
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  PENNIE & EDMONDS
;       STREET:  1155 AVENUE OF THE AMERICAS
;       CITY:  NEW YORK
;       STATE:  NEW YORK
;       COUNTRY:  U.S.A.
;       ZIP:  10036-2711
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/246,583
;       FILING DATE:  20-MAY 1994
;       CLASSIFICATION:  514
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Conlin, Laura A.
;       REGISTRATION NUMBER:  30,742
;       REFERENCE/DOCKET NUMBER:  6923.040
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 790-9090
;       TELEFAX:  (212) 869-8864/9741
;       TELEX:  66141 PENNIE
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  165 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  cDNA
US-08-246-583-4


alignment_scores:
            Quality:    5.00              Length:      5
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_234_242 x US-08-246-583-4     .

 Align seq 1/1  to: US-08-246-583-4  from: 1  to: 165

       4 AspSerGluValAsp 8
         |||||||||||||||
      93 GATAGCGAGGTTGAC 107

seq_name: /tmp2_6/pi.bin/2_iea/5A_COMB.seq:US-08-480-552-16

seq_documentation_block:
; Sequence 16, Application US/08480552
; Patent No. 5665550
;  GENERAL INFORMATION:
;     APPLICANT:  Gudkov, Andrei
;     APPLICANT:  Robinson, Igor B.
;     TITLE OF INVENTION:  Genes And Genetic Elements Associated
;     TITLE OF INVENTION:  With Sensitivity To Chemotherapeutic Drugs
;     NUMBER OF SEQUENCES:  22
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Allegretti & Witcoff, Ltd.
;       STREET:  [illegible] State Street
;       CITY:  Boston
;       STATE:  Massachusetts
;       COUNTRY:  U.S.A.
;       ZIP:  02109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/480,552
;       FILING DATE:  07-JUN-1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/033,086
;       FILING DATE:  09 MAR 1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Keown, Wayne A.
;       REGISTRATION NUMBER:  33,924
```

```
;       REFERENCE/DOCKET NUMBER:  98,454
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617/345 9100
;       TELEFAX:  617/345 9111
;   INFORMATION FOR SEQ ID NO:  15:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  250 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
US 08 480 552 15


alignment scores:
              Quality:     5.00              Length:       5
                Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US 09 126 945B 2 COPY_234_242 x US 08 480 552 15

 Align seq 1/1  to: US 08 480 552 15  from: 1  to: 250

        3 ThrAspSerGluVal 7
          |||||||||||||||
      192 ACGGACTCCGAGGTC 206

Seq name: /cgn, s/ptodata/2/ina/A_CMB_ctp_0  ow 5,5  ,ow 15

seq documentation block:
; Sequence 15, Application US/08480552
; Patent No. 5866404
;  GENERAL INFORMATION:
;     APPLICANT:  Gudkov, Andrei
;     APPLICANT:  Robinson, Igor B.
;     TITLE OF INVENTION:  Genes And Genetic Elements Associated
;     TITLE OF INVENTION:  With Sensitivity To Chemotherapeutic Drugs
;     NUMBER OF SEQUENCES:  22
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Allegretti & Witcoff, Ltd.
;       STREET:  75 State Street
;       CITY:  Boston
;       STATE:  Massachusetts
;       COUNTRY:  U.S.A.
;       ZIP:  02109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/929,208
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/480,552
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Keown, Wayne A.
;       REGISTRATION NUMBER:  33,923
;       REFERENCE/DOCKET NUMBER:  98,454
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617/345 9100
;       TELEFAX:  617/345 9111
;   INFORMATION FOR SEQ ID NO:  16:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  250 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
US 08 929 208 16


alignment scores:
              Quality:     5.00              Length:       5
                Ratio:    1.000                Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US 09 126 945B 2 COPY_234_242 x US 08 929 208 16

 Align seq 1/1  to: US 08 929 208 16  from: 1  to: 250

        3 ThrAspSerGluVal 7
          |||||||||||||||
      192 ACGGACTCCGAGGTC 206

Seq name: /cgn, s/ptodata/2/ina/A_CMB_seq_US_08_905_223_52

seq documentation block:
; Sequence 52, Application US/08905223
; Patent No. 6222029
;  GENERAL INFORMATION:
;     APPLICANT:  Edwards, Jean Baptiste D.
;     APPLICANT:  Imelfort, Aymeric
;     APPLICANT:  Lacroix, Bruno
;     TITLE OF INVENTION:  57 ESTS FOR SECRETED PROTEINS
;     NUMBER OF SEQUENCES:  504
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Knobbe, Martens, Olson & Bear
;       STREET:  501 West Broadway
;       CITY:  San Diego
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  92101-4505
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy Disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  Win95
;       SOFTWARE:  Word
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/905,223
;       FILING DATE:
;       CLASSIFICATION:  536
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Israelsen, Ned A.
;       REGISTRATION NUMBER:  29,655
;       REFERENCE/DOCKET NUMBER:
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (619) 235-8550
;       TELEFAX:  (619) 235-0176
;   INFORMATION FOR SEQ ID NO:  52:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  251 base pairs
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  DOUBLE
;       TOPOLOGY:  LINEAR
;     MOLECULE TYPE:  cDNA
;     ORIGINAL SOURCE:
;       ORGANISM:  Homo Sapiens
;       TISSUE TYPE:  Brain
;     FEATURE:
;       NAME/KEY:  sig_peptide
;       LOCATION:  148..197
;       IDENTIFICATION METHOD:  Von Heijne matrix
;       OTHER INFORMATION:  score 8.4
;       OTHER INFORMATION:  seq LLLICHYLPLSLC/IP
US 08 905 223 52


alignment scores:
              Quality:     5.00              Length:       5
```

```
            Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-126-945B-2_COPY_234_242 x US-08-998-222-52...

 Align seq 1/1  to reverse of: US-08-998-222-52  from: 1  to: 251

        8 ThrAspSerGluVal 7
          |||||||||||||||
      195 ACAGACTCAGAGGTA 181

Seq_name: /tmp_ptodata/_ipa_CA_COMB.seq:US-08-738-367-8

Seq_documentation_block:
; Sequence 8, Application US/08738367
; Patent No. 5827688
;  GENERAL INFORMATION:
;     APPLICANT:  Jacobs, Kenneth
;     APPLICANT:  McCoy, John
;     APPLICANT:  LaVallie, Edward
;     APPLICANT:  Racie, Lisa
;     APPLICANT:  Merberg, David
;     APPLICANT:  Treacy, Maurice
;     APPLICANT:  Spaulding, Vikki
;     TITLE OF INVENTION:  SECRETED PROTEINS AND POLYNUCLEOTIDES
;     TITLE OF INVENTION:  ENCODING THEM
;     NUMBER OF SEQUENCES:  8
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Genetics Institute, Inc.
;       STREET:  87 CambridgePark Drive
;       CITY:  Cambridge
;       STATE:  Massachusetts
;       COUNTRY:  U.S.A.
;       ZIP:  02140
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/738,367
;       FILING DATE:
;       CLASSIFICATION:  514
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Brown, Scott A.
;       REGISTRATION NUMBER:  32,724
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 498-8224
;       TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  8:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  296 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-738-367-8


Alignment_scores:
             Quality:    5.00              Length:       5
               Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment_block:
 US-09-126-945B-2_COPY_234_242 x US-08-738-367-8/rev   ..

 Align seq 1/1  to reverse of: US-08-738-367-8  from: 1  to: 296

        8 ThrAspSerGluVal 7
          |||||||||||||||
       76 ACAGATTCTGAAGTN 62

Seq_name: /tmp_ptodata/_ipa_CA_COMB.seq:US-08-998-416-21

Seq_documentation_block:
; Sequence 21, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;     APPLICANT:  Philippsen, Peter
;     APPLICANT:  Pohlmann, Rainer
;     APPLICANT:  Steiner, Sabine
;     APPLICANT:  Mohr, Christine
;     APPLICANT:  Wendland, Jurgen
;     APPLICANT:  Knechtle, Philipp
;     APPLICANT:  Rebischung, Corinne
;     TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;     TITLE OF INVENTION:  AND USES THEREOF
;     NUMBER OF SEQUENCES:  1152
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  No. 6239264artis Corporation
;       STREET:  3054 Cornwallis Road
;       CITY:  Research Triangle Park
;       STATE:  No. 6239264th Carolina
;       COUNTRY:  USA
;       ZIP:  27709
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/998,416
;       FILING DATE:  24-DEC-1997
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  CH 3016/97
;       FILING DATE:  31-DEC-1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Meiss, J. Timothy
;       REGISTRATION NUMBER:  38,241
;       REFERENCE/DOCKET NUMBER:  33,[illegible]A/CGC1976
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  919 541 8567
;       TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  21:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  441 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     ORIGINAL SOURCE:
;       ORGANISM:  PAG10020
US-08-998-416-21


Alignment_scores:
             Quality:    5.00              Length:       5
               Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment_block:
 US-09-126-945B-2_COPY_234_242 x US-08-998-416-21   ..

 Align seq 1/1  to: US-08-998-416-21  from: 1  to: 441

        5 SerGluValAspSer 9
          |||||||||||||||
      231 TCAGAAGTGGACTCA 245

Seq_name: /tmp_ptodata/_ipa_CA_5990.seq:US-08-664-599B-19

Seq_documentation_block:
; Sequence 19, Application US/08664599B
```

```
; Patent No. 5807703
;  GENERAL INFORMATION:
;     APPLICANT:  Jacobs, Kenneth
;     APPLICANT:  McCoy, John
;     APPLICANT:  LaVallie, Edward
;     APPLICANT:  Racie, Lisa
;     APPLICANT:  Merberg, David
;     APPLICANT:  Treacy, Maurice
;     APPLICANT:  Evans, Cheryl
;     APPLICANT:  Spaulding, Vikki
;     APPLICANT:  Bowman, Michael
;     TITLE OF INVENTION:  SECRETED PROTEINS AND POLYNUCLEOTIDES
;     TITLE OF INVENTION:  ENCODING THEM
;     NUMBER OF SEQUENCES:  47
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Genetics Institute, Inc.
;       STREET:  87 CambridgePark Drive
;       CITY:  Cambridge
;       STATE:  Massachusetts
;       COUNTRY:  U.S.A.
;       ZIP:  02140
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/664,596B
;       FILING DATE:
;       CLASSIFICATION:  514
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Brown, Scott A.
;       REGISTRATION NUMBER:  32,724
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 498 8224
;       TELEFAX:  (617) 876 5851
;   INFORMATION FOR SEQ ID NO:  19:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  454 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US 08 664 596B 19


alignment scores:
            Quality:     5.00              Length:       5
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US 09 126 945B 2 COPY 234 242 x US 08 664 596B 19   ..

 Align seq 1/1  to: US 08 664 596B 19  from: 1  to: 454

        5 SerGluValAspSer 9
          |||||||||||||||
      186 AGTGAGGTTGACAGT 200

seq name: /cgn2 6/ptodata/2/ina/5B COMB.seq.US 08 726 406A 34

seq documentation block:
; Sequence 34, Application US/08726406A
; Patent No. 5958684
;  GENERAL INFORMATION:
;     APPLICANT:  van Leeuwen, Frederik Willem
;     APPLICANT:  Burbach, Johannes Peter Henri
;     APPLICANT:  Grosveld, Franklin G
;     TITLE OF INVENTION:  DIAGNOSIS METHODS AND REAGENTS
;     NUMBER OF SEQUENCES:  189
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Banner & Witcoff, Ltd.
;       STREET:  1 Financial Center
;       CITY:  Boston
;       STATE:  MA
;       COUNTRY:  US
;       ZIP:  02111
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 Mb storage
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  WordPerfect 6.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/726,406A
;       FILING DATE:  02 oct 1996
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  GB 95/20080.4
;       FILING DATE:  02 oct 1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 60/009,842
;       FILING DATE:  01 Jan 1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Williams, Ph.D., Kathleen M.
;       REGISTRATION NUMBER:  34,380
;       REFERENCE/DOCKET NUMBER:  96,348 A (7255/00784)
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 345 9100
;       TELEFAX:  (617) 345 9111
;   INFORMATION FOR SEQ ID NO:  34:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  480 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US 08 726 406A 34


alignment scores:
            Quality:     5.00              Length:       5
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US 09 126 945B 2 COPY 234 242 x US 08 726 406A 34   ..

 Align seq 1/1  to: US 08 726 406A 34  from: 1  to: 480

        1 SerIleThrAspSer 5
          |||||||||||||||
      207 AGCATCACTGACTCC 221

seq name: /cgn2 6/ptodata/2/ina/6A COMB.seq.US 08 961 083 49

seq documentation block:
; Sequence 49, Application US/08961083
; Patent No. 6159469
;  GENERAL INFORMATION:
;     APPLICANT:  Choi et. al.
;     TITLE OF INVENTION:  Streptococcus pneumoniae Antigens and Vaccines
;     NUMBER OF SEQUENCES:  452
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Human Genome Sciences, Inc.
;       STREET:  9410 Key West Avenue
;       CITY:  Rockville
;       STATE:  Maryland
;       COUNTRY:  USA
;       ZIP:  20850
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;       COMPUTER:  HP Vectra 166/33
;       OPERATING SYSTEM:  MS-DOS Version 6.2
;       SOFTWARE:  ASCII Text
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/961,083
```

```
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Brookes, A. Anders
;       REGISTRATION NUMBER:  36,373
;       REFERENCE/DOCKET NUMBER:  P8410P2
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (301) 309-8504
;       TELEFAX:  (301) 309-8512
;   INFORMATION FOR SEQ ID NO:  49:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  490 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
US-08-961-083-49


alignment_scores:
                Quality:     5.00                Length:       5
                  Ratio:  1.000                  Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_234_242 x US-08-961-083-49    ..

 Align seq 1/1  to: US-08-961-083-49  from: 1  to: 490

        5 SerGluValAspSer 9
          |||||||||||||||
      419 ICAGAGGTGGATAGT 433

seq_text  Seq2 % probata_2_ucl_EA_COMB seq:US-08-378-588-20

seq_documentation_block:
; Sequence 20, Application US/08378588
; Patent No. 5608148
;  GENERAL INFORMATION:
;     APPLICANT:  John, Maliyakal E.
;     TITLE OF INVENTION:  TRANSGENIC COTTON PLANTS
;     TITLE OF INVENTION:  PRODUCING HETEROLOGOUS PEROXIDASE
;     NUMBER OF SEQUENCES:  29
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Nicholas J. Seay, Quarles & Brady
;       STREET:  First Wisconsin Plaza, One South
;       STREET:  Pinckney St.,
;       STREET:  P.O. Box 2113
;       CITY:  Madison
;       STATE:  WI
;       COUNTRY:  USA
;       ZIP:  53701-2113
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/378,588
;       FILING DATE:
;       CLASSIFICATION:  800
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Seay, Nicholas J.
;       REGISTRATION NUMBER:  27,486
;       REFERENCE/DOCKET NUMBER:  11 229-9101 2
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (608) 251-2484
;       TELEFAX:  (608) 251-9166
;   INFORMATION FOR SEQ ID NO:  20:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  681 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-378-588-20


alignment_scores:
                Quality:     5.00                Length:       5
                  Ratio:  1.000                  Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_234_242 x US-08-378-588-20    ..

 Align seq 1/1  to: US-08-378-588-20  from: 1  to: 681

        3 IleAspSerGluVal 7
          |||||||||||||||
      639 ATCAGATAGCGAACTA 654
```

```
CM of: US-09-126-945B-2_COPY_234_242 to: EST:*   GCG format : pir

Date: Sep 14, 2001  7:47 PM

About: Results were produced by the GenCore software, version 4.5.
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
  -MODEL frame+_p2n.model -DEV=xlp
  -Q="/cgn2_1/USER/s_specl/US-09-126-945B-2.rst_234_242.seq" -DB="/cgn2_1/pep_pat/pat_1.fil" -DB_EST -QFMT=fastap -SUFFIX=olip2n.rst -GAPOP 4.500
  -GAPEXT 0.050 -MINMATCH=0.100  LOOPCUT=0.000 -LOOPEXT 0.000
  -QGAPOP 4.500 -QGAPEXT 0.050 -XGAPOP=60.000 -XGAPEXT 60.000
  -FGAPOP=6.000  FGAPEXT 7.000 -YGAPOP 60.000 -YGAPEXT 60.000
  -DELOP 6.000 -DELEXT=7.000 -START=1 -MATRIX=olipo
  -TRANS human40.cdi -LIST=45 -DOCALIGN 200 -THR_SCORE=quality
  -THR_MIN=1 -ALIGN 15  MODE LOCAL -OUTPMI pir  NORM ext  MINLEN=
  -MAXLEN=2000000000 -USER US09126945_/CGN1_1_1645 -NCPU 6
  -LOCAL 3 -LONGLOG -NO_XLPXY -WAIT -THREADS 1

Search information block:
Query: US-09-126-945B-2_COPY_234_242
Query length: 9
Database: EST:*
Database sequences: 10228115
Database length: 4314594
Search time (sec): 9.087.630000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP XGAPOP 60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT XGAPEXT 60.000


score_list:
Sequence                Strd Orig    %Score    EScore Len   ! Documentation ..
qb_est78:BE757615         +     9.00  187.39   0.2555   539  ! BE757615 212164 MARC 2BOV Bos t
qb_est92:BG159088         +     8.00  166.65   3.65     549  ! BG159088 RH72_17_F11.b1_A003 R
qb_est27:AJ393331         -     7.00  143.38   72.10    803  ! AJ393331 AJ393331 dkt2426 Gallu
qb_gss5:CNS042E4          +     7.00  141.91   87.16    991  ! AL314149 Tetraodon nigroviridis
qb_gss4:CNS03T8C          -     7.00  141.71   89.44    1020  ! AL269396 Tetraodon nigroviridi
qb_gss4:CNS040HW          -     7.00  141.41   92.97    1065  ! AL269396 Tetraodon nigroviridi
qb_gss5:CNS056IA          +     7.00  141.40   93.12    1067  ! AL022371 Tetraodon nigroviridi
qb_est13:AA930101         -     6.00  139.92   112.52    69  ! AA930101 vq58f12.r1 Barstead mou
qb_est69:BE061497         -     6.00  132.77   281.52    192  ! BE061497 MR0-BT0248-091299-103-
qb_est69:BE061563         -     6.00  132.63   286.27    196  ! BE061563 MR0-BT0248-071099-105-
qb_est68:BB571701         -     6.00  132.35   301.15    207  ! BB571701 BB571701 RIKEN full-l
qb_est20:AI468816         +     6.00  131.66   324.51    225  ! AI468816 ti42b04.x1 NCI_CGAP_Li
qb_est2:AA091451          -     6.00  131.51   330.97    230  ! AA091451 112866.seq.F Human fet
qb_est104:C12830          +     6.00  131.21   343.84    240  ! C12830 C12830 Yuji Kohara unpub
qb_est53:AW885773         -     6.00  130.90   357.92    251  ! AW885773 RC4-CT0071-050500-015-
qb_est105:B91851          -     6.00  130.87   359.20    252  ! B91851 ys81105.s1 Soares retina
qb_est107:R45449          +     6.00  130.71   366.86    258  ! R45449 yq46e11.s1 Soares infant
qb_est75:BE530105         -     6.00  130.63   370.68    261  ! BE530105 M76D14SIM Arabidopsis
qb_est50:AW705589         +     6.00  130.60   371.95    262  ! AW705589 sk50a03.y1 Gm c1019 Gl
qb_est96:BG100351         +     6.00  130.34   384.65    272  ! BG100351 ay05d01.y1 McCarrey EG
qb_est23:AI701657         +     6.00  130.16   393.51    279  ! AI701657 tg98f01.x1 NCI_CGAP_Ki
qb_est51:AW767389         +     6.00  130.14   394.77    280  ! AW767389 da68e02.y1 Rutland str
qb_est84:BE173457         -     6.00  130.14   394.77    280  ! BE173457 MYE1255a Myeloma (MYE)
qb_est62:BB347066         -     6.00  130.01   401.08    285  ! BB347066 BB347066 RIKEN full-le
qb_est106:N54693          -     6.00  129.96   403.60    287  ! N54693 yz10c07.r1 Soares_multip
qb_est92:BF804535         -     6.00  129.89   407.38    290  ! BF804535 QV4-C10155-031100-514-
qb_est66:BB243386         -     6.00  129.84   409.90    292  ! BB243386 BB243386 RIKEN full-le
qb_est8:AV307627          -     6.00  129.58   422.71    304  ! AV307627 AV307627 RIKEN full-le
qb_est80:BF859669         -     6.00  129.52   425.76    307  ! BF859669 wi11f09.x1 NCI_CGAP_Pr
qb_est7:AA401747          +     6.00  129.44   442.47    319  ! AA401747 zl66e07.s1 Soares test
qb_est4:AA292913          -     6.00  129.34   437.47    314  ! AA292913 zl66e07.r1 Soares_test
qb_est26:AI914600         -     6.00  129.16   447.44    322  ! AI914600 wf88e11.x1 NCI_CGAP_CO
qb_est57:BB122930         -     6.00  129.05   453.66    327  ! BB122930 BB122930 RIKEN full-le
qb_est92:BF804673         -     6.00  129.01   456.15    329  ! BF804673 CM0-CT0095-041100-600-
qb_est104:D67854          +     6.00  128.99   457.39    330  ! D67854 CELK0g765r Yuji Kohara u
qb_est69:BE061464         -     6.00  128.88   464.59    335  ! BE061464 MR0-BT0248-261199-101-
qb_est50:AW652623         -     6.00  128.84   466.07    337  ! AW652623 160817 MARC 1b V Bos t
qb_est67:BB543097         -     6.00  128.82   467.21    338  ! BB543097 BB543097 RIKEN full-le
qb_est22:AI572156         -     6.00  128.70   474.23    344  ! AI572156 te37d05.x1 Soares_NhHB
qb_est43:AW145751         -     6.00  128.63   478.45    347  ! AW145751 xn76c01.x1 NCI_CGAP
qb_est18:AI295668         -     6.00  128.40   494.25    359  ! AI295668 IL09459.5prime IL09
qb_est103:C40856          -     6.00  128.38   494.48    360  ! C40856 C40856 Yuji Kohara un
qb_est103:C62171          -     6.00  128.38   494.48    360  ! C62171 C62171 Yuji Kohara un
qb_est108:R85946          -     6.00  128.32   498.17    364  ! R85946 yt66a04.s1 Soares ret
qb_est78:BE719705         -     6.00  128.23   504.31    368  ! BE719705 hC0-BT0862-080800-0

Seq_name: qb_est78:BE757615

Seq_format: GCG_GenBank:
LOCUS       BE757615        539 bp    mRNA            EST       15-SEP-2000
DEFINITION  212164 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BE757615
VERSION     BE757615.1  GI:10171607
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 539)
  AUTHORS   Smith,T.P.L., Casas,E., Stone,R.T., Beaton,M.F., Grosse,W.M.,
            Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Kohrer,G.A., Laegreid
            ,W.W. and Keele,J.W.
  TITLE     Design and use of four pooled tissue normalized cDNA libraries for
            EST discovery in cattle
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            Po Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRIMERS
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 64  row: G  column: 13
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1..539
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 2BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from testis, thymus,
                     semitendonosus muscle, longissimus muscle, pancreas,
                     adrenal, and endometrium."
BASE COUNT        98 a    168 c    193 g     80 t
ORIGIN


alignment_scores:
             quality:    9.00              Length:      9
               Ratio:   1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2.rst_234_242 x BE757615    .

 Align_seq 1/1  to: BE757615  from: 1  to: 539

       1 SerTrpThrAspSerGlnValAspSer 9
         |||| |||||||||||||||||||||
     452 AGCTGGACGGACAGCGAGGTGGACTCG 478

Seq_name: qb_est92:BG159088

Seq_documentation_block:
```

```
LOCUS       BG159088       549 bp    mRNA            EST       06-FEB-2001
DEFINITION  RHIZ2_17_F11.b1_A003 Rhizome2 (RHIZ2) Sorghum propinquum cDNA, mRNA
            sequence.
ACCESSION   BG159088
VERSION     BG159088.1  GI:12692772
KEYWORDS    EST.
SOURCE      Sorghum propinquum.
  ORGANISM  Sorghum propinquum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 549)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: Sorghum propinquum rhizomes
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Seq primer: -40M-REV
            High quality sequence stop: 486
            POLYA No.
FEATURES             Location/Qualifiers
     source          1..549
                     /organism="Sorghum propinquum"
                     /db_xref="taxon:132711"
                     /clone_lib="Rhizome2 (RHIZ2)"
                     /note="Organ: Rhizomes; Vector: pBluescript II from Lambda
                     Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                     from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      130 a    133 c    156 g    130 t
ORIGIN


alignment_scores:
           Quality:     8.00              Length:       8
             Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_234_242 x BG159088 ..

 Align seq 1/1  to: BG159088  from: 1  to: 549

       2 TrpThrAspSerGluValAspSer 9
         ||||||||||||||||||||||||
     520 TGGACTGACAGTGAAGTTGACTCC 543

seq_name: gb_est27:AJ393331

seq_documentation_block:
LOCUS       AJ393331       803 bp    mRNA            EST       25-JAN-2001
DEFINITION  AJ393331 dkfz426 Gallus gallus cDNA clone 15e3r1, mRNA sequence.
ACCESSION   AJ393331
VERSION     AJ393331.1  GI:7122087
KEYWORDS    EST.
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 803)
  AUTHORS   Abdrakhmanov,I., Lodygin,D., Geroth,P., Arakawa,H., Law,A., Plachy
            ,J., Korn,B. and Buerstedde,J.M.
  TITLE     A large database of chicken bursal ESTs as a resource for the
            analysis of vertebrate gene function
  JOURNAL   Genome Res. 10 (12), 2062-2069 (2000)
  MEDLINE   20568495
COMMENT     Contact: Buerstedde JM
            Cellular Immunology
            Heinrich-Pette-Institute
            Martinistr. 52, 20251 Hamburg, Germany
            Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
FEATURES             Location/Qualifiers
     source          1..803
                     /organism="Gallus gallus"
                     /strain="CB"
                     /db_xref="taxon:9031"
                     /clone="15e3r1"
                     /clone_lib="dkfz426"
                     /tissue_type="Bursa of Fabricius"
BASE COUNT      265 a    155 c    186 g    194 t      3 others
ORIGIN


alignment_scores:
           Quality:     7.00              Length:       7
             Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_234_242 x AJ393331/rev ..

 Align seq 1/1  to reverse of: AJ393331  from: 1  to: 803

       3 ThrAspSerGluValAspSer 9
         |||| ||||||||||||||||
     547 ACAGATTCGGAAGTTGACTCC 567

seq_name: gb_gss5:CNS04ZE4

seq_documentation_block:
LOCUS       CNS04ZE4       991 bp    DNA             GSS       26-JUL-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            043A20 of library B from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL314149
VERSION     AL314149.1  GI:9547033
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 991)
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
REFERENCE   2  (bases 1 to 991)
  AUTHORS   Crollius,H.R., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C.,
            Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W.,
            Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
REFERENCE   3  (bases 1 to 991)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/tetraodon.
FEATURES             Location/Qualifiers
     source          1..991
```

```
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="043A20"
                     /clone_lib="B"
                     /note="Genoscope sequence ID : CGAB043BA10C1 end : 17"
BASE COUNT       401 a    227 c    185 g    248 t     30 others
ORIGIN


alignment_scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_234_242 x CNS04ZE4    .

 Align seq 1/1  to: CNS04ZE4  from: 1  to: 991

       1 SerIrpThrAspSerGluVal 7
         |||||||||||||||||||||
     728 AGCTGGACAGACTCAGAAGTG 748

Seq_name: gb_gss4:CNS043T8C

Seq_documentation_block:
LOCUS        CNS043T8C     1020 bp    DNA              GSS       17-MAY-2000
DEFINITION   Tetraodon nigroviridis genome survey sequence PUC-ori end of clone
             056A16 of library G from Tetraodon nigroviridis, genomic survey
             sequence.
ACCESSION    AL259509
VERSION      AL259509.1  GI:7988521
KEYWORDS     GSS; genome survey sequence.
SOURCE       Tetraodon nigroviridis.
  ORGANISM   Tetraodon nigroviridis
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
             Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
             Tetraodontidae; Tetraodon.
REFERENCE    1  (bases 1 to 1020)
  AUTHORS    Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
             Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
             Weissenbach,J.
  TITLE      Characterization and repeat analysis of the compact genome of the
             freshwater pufferfish Tetraodon nigroviridis
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 1020)
  AUTHORS    Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
             Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
             Saurin,W. and Weissenbach,J.
  TITLE      Human gene number estimate provided by genome wide analysis using
             Tetraodon nigroviridis DNA sequence
  JOURNAL    Unpublished
REFERENCE    3  (bases 1 to 1020)
  AUTHORS    Genoscope.
  TITLE      Direct Submission
  JOURNAL    Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases.
COMMENT      This sequence is a single read and was generated as part of a large
             scale clone-end sequencing project of the Tetraodon nigroviridis
             genome. For more information, please take a look at
             http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1..1020
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="056A16"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : CGBG056BA08SP1 end :
                     PUC-ori"
BASE COUNT       269 a    217 c    295 g    239 t     % others
ORIGIN
```

```
alignment_scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_234_242 x CNS043T8C/rev    ..

 Align seq 1/1  to reverse of: CNS043T8C  from: 1  to: 1020

       1 SerIrpThrAspSerGluVal 7
         |||||||||||||||||||||
     649 AGCTGGACAGACTCAGAAGTC 629

Seq_name: gb_gss4:CNS0400UW

Seq_documentation_block:
LOCUS        CNS0400UW     1065 bp    DNA              GSS       18-MAY-2000
DEFINITION   Tetraodon nigroviridis genome survey sequence 17 end of clone
             043A20 of library G from Tetraodon nigroviridis, genomic survey
             sequence.
ACCESSION    AL269393
VERSION      AL269393.1  GI:7991279
KEYWORDS     GSS; genome survey sequence.
SOURCE       Tetraodon nigroviridis.
  ORGANISM   Tetraodon nigroviridis
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
             Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
             Tetraodontidae; Tetraodon.
REFERENCE    1  (bases 1 to 1065)
  AUTHORS    Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
             Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
             Weissenbach,J.
  TITLE      Characterization and repeat analysis of the compact genome of the
             freshwater pufferfish Tetraodon nigroviridis
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 1065)
  AUTHORS    Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
             Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
             Saurin,W. and Weissenbach,J.
  TITLE      Human gene number estimate provided by genome wide analysis using
             Tetraodon nigroviridis DNA sequence
  JOURNAL    Unpublished
REFERENCE    3  (bases 1 to 1065)
  AUTHORS    Genoscope.
  TITLE      Direct Submission
  JOURNAL    Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases.
COMMENT      This sequence is a single read and was generated as part of a large
             scale clone-end sequencing project of the Tetraodon nigroviridis
             genome. For more information, please take a look at
             http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1..1065
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="043A20"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : CGBG056BA10C1 end : 17"
BASE COUNT       267 a    254 c    267 g    272 t     5 others
ORIGIN


alignment_scores:
            Quality:    7.00              Length:       7
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_234_242 x CNS0400UW/rev    ..

 Align seq 1/1  to reverse of: CNS0400UW  from: 1  to: 1065
```

```
       1 SerTrpThrAspSerGluVal 7
         |||||||||||||||||||||
     450 AGCTGGACAGATTCTGAAGTC 430

Seq name: qb qss5:CNS0561A

Seq documentation block:
LOCUS       CNS0561A      1067 bp    DNA             GSS       26 JUL 2000
DEFINITION  Tetraodon nigroviridis genome survey sequence 1/ end of clone
            025G05 of library B from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL423371
VERSION     AL423371.1  GI:3756254
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1067)
  AUTHORS   Roest Crollius,H., Jaillon,O., Bastiya,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (3), 235-238 (2000)
  MEDLINE   20296633
REFERENCE   2  (bases 1 to 1067)
  AUTHORS   Crollius,H.R., Jaillon,O., Bastiya,C., Ozouf-Costaz,C., Fizames,C.,
            Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W.,
            Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
REFERENCE   3  (bases 1 to 1067)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12 APR 2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a simple read and was generated as part of a large
            scale clone end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/tetraodon.
FEATURES             Location/Qualifiers
     source          1..1067
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="025G05"
                     /clone_lib="B"
                     /note "Genoscope sequence ID : C0AB025Ab03C1 end : F7"
BASE COUNT       324 a    245 c    200 g    281 t     17 others
ORIGIN


Alignment scores:
            Quality:    7.00              Length:       7
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-126-945B-2-COPY_234_242 x CNS0561A  ..

 Align seq 1/1  to: CNS0561A  from: 1  to: 1067

       1 SerTrpThrAspSerGluVal 7
         |||||||||||||||||||||
     790 AGCTGGACAGACTCAGAAGTG 810

Seq name: qb est13:AA930101

Seq documentation block:
LOCUS       AA930101       69 bp    mRNA            EST       23 APR 1998
```

```
DEFINITION  zy01l2.r1 Barstead mouse proximal colon MPLRB6 Mus musculus cDNA
            clone IMAGE:1106543 5', similar to gb:X80159 M.musculus CW17 mRNA
            (MOUSE);, mRNA sequence.
ACCESSION   AA930101
VERSION     AA930101.1  GI:3079694
KEYWORDS    EST
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 69)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Stoptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU HBMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HBMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, Mo 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:604711
            Trace considered overall poor quality
            Seq primer: -20m13 rev2 EI from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1..69
                     /organism="Mus musculus"
                     /strain="BALB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1106543"
                     /clone_lib="Barstead mouse proximal colon MPLRB6"
                     /dev_stage="7 day juvenile"
                     /lab_host="DH10B"
                     /note "Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACGAATTCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTT
                     3'];  double-stranded cDNA was ligated to Eco RI adaptors
                     [AATTCCGGATCCTTGT], digested with Not I and cloned into the
                     Not I and Eco RI sites of  the modified pT7T3 vector.
                     Library constructed by Bob Barstead.  "
BASE COUNT          9 a     19 c     18 g     23 t
ORIGIN


Alignment scores:
            Quality:    6.00              Length:       6
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-126-945B-2-COPY_234_242 x AA930101  ..

 Align seq 1/1  to reverse of: AA930101  from: 1  to: 69

       1 SerTrpThrAspSerGlu 6
         ||||||||||||||||||
      45 TCTGGACCAGACTCTGAG 18

Seq name: qb est69:BE061497

Seq documentation block:
LOCUS       BE061497       192 bp    mRNA            EST       09 JUN 2000
DEFINITION  MR0-BI0248-091299-103-g08-BI0248 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE061497
VERSION     BE061497.1  GI:8406147
```

```
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 192)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brumstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/cgi-bin/db_hits.exe?qtype=est&qval=
            299-103-q08&t3=1999-12-09&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 191
FEATURES             Location/Qualifiers
     source          1..192
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib "RT0248"
                     /dev_stage "Adult"
                     /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT       56 a      50 c      41 g      55 t
ORIGIN


alignment_scores:
            Quality:    6.00             Length:      6
              Ratio:   1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_234_242 x BE061563/rev   ..

 Align seq 1/1  to reverse of: BE061563  from: 1  to: 192

        1 SerTrpThrAspSerGlu 6
          ||||||||||||||||||
      55 AGTTGGACAGACTCTGAA 38

seq_name: db_est69:BE061563

seq_documentation_block:
LOCUS       BE061563      196 bp    mRNA            EST       09-JUN-2000
DEFINITION  MR0-BT0248-070100-105-f07 BT0248 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE061563
VERSION     BE061563.1  GI:8406213
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 196)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brumstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/cgi-bin/db_hits.exe?qtype=est&qval=
            100-105-107&t3=2000-01-07&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 7
            High quality sequence stop: 187.
FEATURES             Location/Qualifiers
     source          1..196
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib "RT0248"
                     /dev_stage "Adult"
                     /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT       56 a      52 c      26 g      59 t       3 others
ORIGIN


alignment_scores:
            Quality:    6.00             Length:      6
              Ratio:   1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_234_242 x BE061563/rev   ..

 Align seq 1/1  to reverse of: BE061563  from: 1  to: 196

        1 SerTrpThrAspSerGlu 6
          ||||||||||||||||||
      55 AGTTGGACAGACTCTGAA 38

seq_name: db_est68:BB571701

seq_documentation_block:
LOCUS       BB571701      207 bp    mRNA            EST       30-NOV-2000
DEFINITION  BB571701 RIKEN full-length enriched, 0 day neonate head Mus
            musculus cDNA clone 4831401L05 5', mRNA sequence.
ACCESSION   BB571701
VERSION     BB571701.1  GI:11468245
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 207)
  AUTHORS   Akiyama,F., Akahira,S., Akimura,I., Arai,A., Arakawa,I., Carninci,P.
            , Hanagaki,T., Hayatsu,N., Hiraoka,T., Hirozane,I., Hodoyama,Y.,
```

```
                Imotani,K., Ishii,Y., Itoh,M., Itoo,M., Kawai,J., Kojima,S., Kono
                ,H., Muramatsu,M., Matsuyama,T., Miyazaki,A., Nishimura,M., Nishi,K.,
                Nomura,K., Numazaki,R., Okazaki,Y., Okido,T., Osa,A., Saka,C.,
                Sakai,K., Sasaki,D., Sato,K., Shibata,K., Shibata,Y., Shinagawa,A.,
                Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Tanaka
                ,T., Toya,T., Watahiki,A., Yamamura,T., Yasunishi,A., Yoshida,K.,
                Yoshiki,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE         RIKEN Mouse ESTs (Aizawa,K. et al. 2000)
  JOURNAL       Unpublished (2000)
COMMENT         Contact: Yoshihide Hayashizaki
                Genome Exploration Research Group, Life Science Tsukuba Center,
                Genome Science Laboratory
                The Institute of Physical and Chemical Research (RIKEN), Genomic
                Sciences Center
                3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
                Tel: +81 298-36-9013
                Fax: +81 298-36-9098
                Email: genome res@rtc.riken.go.jp,
                URL:http://genome.rtc.riken.go.jp/
                Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
                ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
                Thermostabilization and thermoactivation of thermolabile enzymes by
                trehalose and its application for the synthesis of full length
                cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
                Itoh,M., Kitsunai,T., Akiyama,T., Shibata,K., Izawa,M., Kawai,J.,
                Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
                ,Y. and Hayashizaki,Y.
                Automated titration-based high throughput plasmid preparation
                system. Genome Res. 9 (5), 463-470 (1999)
                Carninci,P. and Hayashizaki,Y.
                High efficiency full-length cDNA cloning. Methods Enzymol. 303,
                19-44 (1999)
                Please visit our web site (http://genome.rtc.riken.go.jp) for
                further details.
FEATURES                Location/Qualifiers
     source             1..207
                        /organism="Mus musculus"
                        /strain="C57BL/6J"
                        /db_xref="taxon:10090"
                        /clone="4831431I05"
                        /clone_lib="RIKEN full-length enriched, 0 day neonate
                        head"
                        /sex="mixed"
                        /tissue_type="head"
                        /dev_stage="0 day neonate"
                        /lab_host="DH10B"
                        /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                        prepared and sequenced in Mouse Genome Encyclopedia
                        Project of Genome Exploration Research Group in Riken
                        Genomic Sciences Center and Genome Science Laboratory in
                        RIKEN. Division of Experimental Animal Research in Riken
                        contributed to prepare mouse tissues. 1st strand cDNA was
                        primed with a primer [5'
                        GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                        prepared by using trehalose thermo-activated reverse
                        transcriptase and subsequently enriched for full-length by
                        cap trapper. cDNA went through one round of normalization
                        to Rot=10.0 and subtraction to Rot=100.0. Second
                        strand cDNA was prepared with the primer adapter of
                        sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC
                        3']. cDNA was cloned into the XhoI and BamHI sites.
                        Vector: a modified pBluescript KS(+) after bulk excision
                        from Lambda ZAP 1."
BASE COUNT        55 a      46 c      54 g      52 t
ORIGIN


alignment scores:
            Quality:    6.00              Length:      6
              Ratio:   1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-126-945B-2_COPY_234_242 x AI468816  ..

 Alignment 1/1  1 : AI468816  from: 1  to: 225

       2 IrpThrAspSerGluVal 7
         ||||||||||||||||||
     128 IGGACGGACTCTGGAAGTT 111

Seq_name: gb_est20:AI468816

Seq_documentation_block:
LOCUS         AI468816       225 bp    mRNA             EST       30-MAR-1999
DEFINITION  (clone 2133103) NCI_CGAP_Lym12 Homo sapiens cDNA clone IMAGE:2133103 3'
            similar to contains L1.t1 L1 repetitive element ;, mRNA sequence.
ACCESSION   AI468816
VERSION     AI468816.1  GI:4330906
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 225)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Life Technologies catalog #: 11547-015
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www.bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1181  Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 214.
FEATURES                Location/Qualifiers
     source             1..225
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="IMAGE:2133103"
                        /clone_lib="NCI_CGAP_Lym12"
                        /tissue_type="lymphoma, follicular mixed small and large
                        cell"
                        /lab_host="DH10B"
                        /note="Organ: lymph node; Vector: pCMV-SPORT6; Site_1:
                        SalI; Site_2: NotI; Cloned unidirectionally.  Primer:
                        oligo dT.  Average insert size 1.25 kb. Life Technologies
                        catalog #: 11547-015"
BASE COUNT        95 a      38 c      34 g      58 t
ORIGIN


alignment scores:
            Quality:    6.00              Length:      6
              Ratio:   1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-126-945B-2_COPY_234_242 x AI468816  ..

 Alignment 1/1  1 : AI468816  from: 1  to: 225

       2 IrpThrAspSerGluVal 7
         ||||||||||||||||||
      86 IGGACTGATAGTGGAAGTT 103

Seq_name: gb_est2:AA091451

Seq_documentation_block:
LOCUS         AA091451       230 bp    mRNA             EST       24-OCT-1996
DEFINITION  tl2866.seq.F Human fetal heart, Lambda ZAP Express Homo sapiens
            cDNA 5', mRNA sequence.
```

```
ACCESSION   AA091451
VERSION     AA091451.1  GI:1636906
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 230)
  AUTHORS   Liew,C.C.
  TITLE     cDNAs from fetal heart (1996)
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Liew CC
            Department of Laboratory Medicine and Pathobiology
            University of Toronto
            Banting Institute, 100 College St., Toronto, Ontario, M5G1L5
            Tel: 4169788758
            Fax: 4169785650
            Email: liewcc@utcc.utoronto.ca
            PCR Primers
            FORWARD: 5' GCCAAGCTCGAAATTAACCCTCACTAAAGGG 3'
            BACKWARD: 5' CCAGTGAATTGTAATACGACTCACTATAGGGCG 3'
            Seq primer: 5' GAAATTAACCCTCACTAAAGGG 3'.
FEATURES             Location/Qualifiers
     source          1..230
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib "Human fetal heart, Lambda ZAP Express"
                     /lab_host="E. coli XL1-Blue"
                     /note="Vector: Lambda ZAP Express; Site_1: EcoRI; Site_2:
                     XhoI; mRNA was purified from human fetal hearts (8-10
                     weeks). cDNA was synthesized using a XhoI-oligo dT
                     adaptor-primer. EcoRI adaptors were ligated, followed by
                     digestion with XhoI, for directional cloning into
                     predigested lambda ZAP Express."
BASE COUNT       60 a      41 c      45 g      84 t
ORIGIN


alignment_scores:
            Quality:    6.00            Length:      6
              Ratio:   1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_234_242 x AA091451/rev

 Align seq 1/1  to reverse of AA091451  from: 1  to: 240

       4 AspSerGluValAspSer 9
         |||||||||||||||||||
     220 GACTCTGAAGTTGACTCA 203

seq_name: qb_est103:C12830

seq_documentation_block:
LOCUS       C12830        240 bp    mRNA            EST       28-DEC-1998
DEFINITION  C12830 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
            clone yk159q10 5', mRNA sequence.
ACCESSION   C12830
VERSION     C12830.1  GI:1560383
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
            ; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 240)
  AUTHORS   Kohara,Y., Motohashi,I., Tabara,H., Watanabe,H., Sugimoto,A., Sano
            ,M., Miyata,A. and Nishigaki,A.
  TITLE     Expression map of the C.elegans genome
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81 559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1..240
                     /organism="Caenorhabditis elegans"
                     /strain "CB1489 him-8(e1489)"
                     /db_xref "taxon:6239"
                     /clone "yk159q10"
                     /clone_lib "Yuji Kohara unpublished cDNA"
                     /sex "hermaphrodite, male"
                     /tissue_type="whole animal"
                     /dev_stage="varied"
BASE COUNT       75 a      44 c      51 g      65 t       6 others
ORIGIN


alignment_scores:
            Quality:    6.00            Length:      6
              Ratio:   1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_234_242 x C12830  ..

 Align seq 1/1  to: C12830  from: 1  to: 240

       8 ThrAspSerGluValAsp 8
         ||||||||||||||||||
     151 ACTGATTCTGAAGTTGAT 168

seq_name: qb_est53:AW885773

seq_documentation_block:
LOCUS       AW885773        251 bp    mRNA            EST       23-MAY-2000
DEFINITION  RC4-OT0071-050500-015-c12 OT0071 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW885773
VERSION     AW885773.1  GI:8047785
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 251)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brumstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            http://www.ludwig.org.br/cgi-bin/clones/getbest12.pl?id=RC4-OT0071-050
            500-015-c12&t3=2000-05-05&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 251.
FEATURES             Location/Qualifiers
     source          1..251
                     /organism="Homo sapiens"
                     /db_xref "taxon:9606"
```

/clone_lib "OT0071"
/dev_stage "Adult"
/note "organ: ovary; Vector: pucl8; Site 1: SmaI; Site 2: SmaI; A mini library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 67 a 48 c 64 g 72 t
ORIGIN

Alignment scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

Alignment block:
US-09-126-945B-2_copy_234_242 x AW885773/rev ..

Align seq 1/1 to reverse of: AW885773 from: 1 to: 251

4 AspSerGluValAspSer 9
|||||||||||||||||||
56 GACAGTGAGGTCGATTCT 39